

# Random linear under-determined systems with block-sparse solutions – asymptotics, large deviations, and finite dimensions

MIHAILO STOJNIC \*

## Abstract

In this paper we consider random linear under-determined systems with block-sparse solutions. A standard subvariant of such systems, namely, precisely the same type of systems without additional block structuring requirement, gained a lot of popularity over the last decade. This is of course in first place due to the success in mathematical characterization of an  $\ell_1$  optimization technique typically used for solving such systems, initially achieved in [4, 12] and later on perfected in [9, 10, 31, 32]. The success that we achieved in [31, 32] characterizing the standard sparse solutions systems, we were then able to replicate in a sequence of papers [26, 30, 34, 35] where instead of the standard  $\ell_1$  optimization we utilized its an  $\ell_2/\ell_1$  variant as a better fit for systems with block-sparse solutions. All of these results finally settled the so-called threshold/phase transitions phenomena (which naturally assume the asymptotic/large dimensional scenario). Here, in addition to a few novel asymptotic considerations, we also try to raise the level a bit, step a bit away from the asymptotics, and consider the finite dimensions scenarios as well.

**Index Terms:** Linear systems of equations; block-sparse solutions;  $\ell_2/\ell_1$ -optimization.

## 1 Introduction

We will start by introducing key mathematical structures that we will need in the rest of the paper. However, we do emphasize right here at the beginning that throughout the presentation we will often assume a high level of familiarity with many well known concepts (we will however try to maintain as much consistency with [26, 30, 34, 35] as possible so that one can consult these earlier works and results obtained therein without much need for additional adjustments).

Let system matrix  $A$  be an  $M \times N$  ( $M \leq N$ ) dimensional matrix with real entries and let  $\tilde{\mathbf{x}}$  be an  $N$  dimensional vector that also has real entries. Additionally, let  $\tilde{\mathbf{x}}$  have no more than  $K$  nonzero entries (one usually calls such a vector  $K$ -sparse). Then one forms the product of  $A$  and  $\tilde{\mathbf{x}}$  to obtain  $\mathbf{y}$

$$\mathbf{y} = A\tilde{\mathbf{x}}. \quad (1)$$

The problems of interest in this paper (and in general in linear systems known to have sparse solutions) are essentially an inverted version of (1). Namely, given  $A$  and  $\mathbf{y}$  from (1), can one then determine  $\tilde{\mathbf{x}}$ ? Mathematically, one asks for the  $k$  sparse solution of

$$A\mathbf{x} = \mathbf{y}, \quad (2)$$

knowing of course (based on (1)) that such a solution exists (for  $K < M/2$  it is in fact unique; we also additionally assume that there is no  $\mathbf{x}$  in (2) that is less than  $K$  sparse). Alternatively, one often rewrites the problem described above in the following way

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_0 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}, \end{aligned} \quad (3)$$

---

\*e-mail: flatoyer@gmail.com

where  $\|\mathbf{x}\|_0$  is what is typically called  $\ell_0$  norm of vector  $\mathbf{x}$  (for all practical purposes needed here, we will think of  $\|\mathbf{x}\|_0$  as being a number that is equal to the number of nonzero entries of  $\mathbf{x}$ ).

Now that we described the problem one then wonders how it can be solved. Quite a few very successful algorithms have been developed over last several decades (see, e.g. [8, 11, 13, 21, 22, 40, 42]). In our view however, the most important one is the following linear programming relaxation of (3), called  $\ell_1$ -optimization

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{y}. \end{aligned} \tag{4}$$

Its importance is of course rooted in its excellent performance characteristics. However, we do believe that its performance characterizations initially done in [4, 12] and later on perfected in [9, 10, 31, 32] also contributed to a large degree to its an overall popularity. In fact, we believe that [4, 12] in particular generated a substantial portion of the interest in linear systems over the last decade. As the main concern of this paper is a bit more specific type of linear systems we will stop short of reviewing in detail results that relate to (2) and to (4). Instead we will refer to [9, 10, 31, 32] and in particular to [29] where a whole lot more can be found.

As mentioned above, in this paper, we will be interested in a bit more specific version of the problem from (2). Namely, we will consider the linear systems with solutions that are block-sparse (more on systems with these types of solutions and their potential applications and recovery algorithms can be found in a series of recent references, see e.g. [1, 5, 14–16, 18, 24, 34, 35, 38] as well as e.g. [2, 3, 6, 7, 17, 20, 23, 39, 41, 43–46] and many references therein for a related problem of recovering jointly sparse vectors). To facilitate the description of the block-sparse vectors we will assume that integers  $N$  and  $d$  are chosen such that  $n = \frac{N}{d}$  is an integer and it represents the total number of blocks that  $\mathbf{x}$  consists of. Clearly  $d$  is the length of each block (typically we will assume  $d \geq 2$  to distinguish between the standard version of the problem from (2)). Furthermore, we will assume that  $\mathbf{X}_i = \mathbf{x}_{(i-1)d+1:id}$ ,  $1 \leq i \leq n$ , are the  $n$  blocks of  $\mathbf{x}$ . One can sometimes also assume that  $m = \frac{M}{d}$  is an integer as well; however, we will specifically emphasize if/when we do that. We will call any vector  $\mathbf{x}$   $k$ -block-sparse if its at most  $k = \frac{K}{d}$  blocks  $\mathbf{X}_i$  are non-zero (non-zero block is a block that is not a zero block; zero block is a block that has all elements equal to zero). Since  $k$ -block-sparse signals are  $K$ -sparse one could then use (4) to recover the solution of (2). Moreover, all of known results related to performance characterization of (4) would translate immediately as well. However, using (4) to recover block-sparse vectors would utilized their block structure in no way. Since this knowledge is a priori available one would be tempted to believe that it could be somehow utilized in the design of the recovery algorithms so that they outperform the standard  $\ell_1$  from (4). This was, of course observed, long time ago and there are of course many ways how one can attempt to exploit the block-sparse structure. Here we will focus on the following polynomial-time algorithm (essentially a combination of  $\ell_2$  and  $\ell_1$  optimizations) that was considered in [38] (see also e.g. [1, 15, 43, 45, 46])

$$\begin{aligned} \min \quad & \sum_{i=1}^n \|\mathbf{x}_{(i-1)d+1:id}\|_2 \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{y}. \end{aligned} \tag{5}$$

Various aspects of the performance analysis of the optimization problem from (5) will be the main subject of the remaining sections of this paper. We will start by reviewing a few known asymptotic results related to the performance characterization of (5). We will then present a few novel asymptotic considerations and will finish things off by providing a collection of results in a non-asymptotic setup.

## 2 Asymptotics

In [38] we started studying (5) as a tool for solving (2). We conducted a serious of numerical experiments as well as a preliminary theoretical analysis of certain properties of (5). Through numerical experiments we demonstrated that as  $d$  grows the algorithm in (5) significantly outperforms the standard  $\ell_1$ . Of course, as hinted above, just based on the structure of (5) it was to be expected that it will be working better than just plain simple  $\ell_1$  from (4). What was perhaps a bit surprising (and very welcomed) was that the improvement over  $\ell_1$  was rather substantial. In fact, this was particularly true in the so-called under-sampled scenario

(i.e. when  $m \ll n$ ) where for fairly moderate values of block-length  $d$  (5) was already approaching the theoretically limiting level of performance.

Providing a rigorous mathematical justification for these observations at the time was completely out of reach. Nonetheless, we were able to show the following: let  $A$  be an  $M \times N$  matrix with a basis of null-space comprised of i.i.d. Gaussian elements; if  $\alpha = \frac{M}{N} \rightarrow 1$  as  $N \rightarrow \infty$  then there is a constant  $d$  such that all  $k$ -block-sparse signals  $\mathbf{x}$  with sparsity  $K \leq \beta N$ ,  $\beta \rightarrow \frac{1}{2}$ , can be recovered with overwhelming probability by solving (5). The precise relation between  $d$  and how fast  $\alpha \rightarrow 1$  and  $\beta \rightarrow \frac{1}{2}$  was quantified in [38] as well. The result was obtained for a fairly narrow range of problem parameters but was in its nature optimal. We of course then recognized the overall potential of (5) and undertook a thorough and systematic study of its performance characteristics. In [34, 35] we extended the results from [38] and obtained the values of the recoverable block-sparsity for any  $\alpha$ , i.e. for  $0 \leq \alpha \leq 1$ . More precisely, for any given constant  $0 \leq \alpha \leq 1$  we in [34, 35] determined a constant  $\beta = \frac{K}{N}$  such that for a sufficiently large  $d$  (5) with overwhelming probability recovers any  $k$ -block-sparse vector of sparsity less than  $K$  (in this paper we, as usual, under overwhelming probability assume a probability that is no more than a number exponentially decaying in  $N$  away from 1).

For algorithms that exhibit the so-called phase-transition (PT) phenomenon, for any given constant  $\alpha \leq 1$  there is a maximum allowable value of  $\beta$  such that for *any* given  $K$ -sparse  $\mathbf{x}$  in (2) the solution that the algorithm offers is with overwhelming probability exactly that given  $K$ -sparse  $\mathbf{x}$ . To be a bit more precise and in an alignment with some of our earlier works, we should add that this type of phase transitions is often called the *strong* phase transition (the *strong* PT) and the value of  $\beta$  that we just described is typically referred to as the *strong threshold* (see [10, 26]). These threshold values are essentially the points (proportions of the system dimensions) where the algorithms (in our case here (5)) exhibit the phase-transition phenomenon. In a bit less formal language, the phase-transition phenomenon essentially means that if the problem dimensions are such that the pair  $(\alpha, \beta)$  is below the so called phase-transition curve (the PT curve) then the algorithm (here (5)) solves the problem (here (2)); otherwise it fails. The ultimate goal of an asymptotic analysis of any algorithm with phase transition is to determine the PT curve.

We should also mention that the above requirement is, from a practical point of view, a bit restrictive. Namely, insisting that the algorithm succeeds for *any* given  $K$ -sparse  $\mathbf{x}$  is a bit too much to ask for if one only cares about a “typical” level of performance. To characterize a bit more typical levels of performance one often relaxes the above PT description. Namely, for any given constant  $\alpha \leq 1$  and *any* given  $\mathbf{x}$  with a given fixed location and a given fixed directions of non-zero blocks there will be a maximum allowable value of  $\beta$  such that (5) finds that given  $\mathbf{x}$  in (2) with overwhelming probability. Such a value of  $\beta$ , which we will here denote by  $\beta_w$ , is typically called the *weak threshold* and the resulting curve the *weak* phase-transition curve (see, e.g. [32, 33]). More important than the name itself is that this type of phase-transition is designed to capture a typical performance in a better way. Indeed, when one needs (2) solved, it typically wants it solved for an  $\mathbf{x}$  or for a set of  $\mathbf{x}$  but quite likely not for every single  $\mathbf{x}$ . In scenarios when this is indeed true the above weak phase-transition curve is highly likely to be more useful.

Now, returning back to what was done in [34, 35], we should emphasize another subtle point. Namely, [34, 35] provided fairly sharp strong threshold values but they had done so in a so to say block asymptotic sense, i.e. the analyses presented in [34, 35] assumed fairly large values of the block-length  $d$ . As such they then provided an ultimate performance limit of  $\ell_2/\ell_1$ -optimization rather than its performance characterization as a function of a particular fixed block-length.

Finally in [26] we were able to massively extended the results of [34, 35] while moving away from the block-asymptotic regime. We essentially introduced a novel probabilistic framework for performance characterization of (5) through which we were finally able to view block-length as a parameter of the system (some of the key components of the framework were actually introduced in [32]). Through the framework we were able to lower-bound  $\beta_w$ . Moreover, the obtained lower bounds were in an excellent agreement with the values obtained for  $\beta_w$  through numerical simulations. In [30] we then showed that the lower bounds on  $\beta_w$  obtained in [26] are actually exact. We below recall on a theorem that essentially summarizes the results obtained in [26, 30] and effectively establishes for any  $0 < \alpha \leq 1$  the exact value of  $\beta_w$  for which (5) finds the  $k$ -block-sparse  $\mathbf{x}$  from (2).

**Theorem 1.** ([26, 30] *Exact weak threshold; block-sparse  $\mathbf{x}$* ) *Let  $A$  be an  $M \times N$  matrix in (2) with i.i.d. standard normal components. Let the unknown  $\mathbf{x}$  in (2) be  $k$ -block-sparse with the length of its blocks  $d$ . Further, let the location and the directions of nonzero blocks of  $\mathbf{x}$  be arbitrarily chosen but fixed. Let  $k, m, n$  be large and let  $\alpha = \frac{m}{n}$  and  $\beta_w = \frac{k}{n}$  be constants independent of  $m$  and  $n$ . Let  $\gamma_{inc}(\cdot, \cdot)$  and  $\gamma_{inc}^{-1}(\cdot, \cdot)$  be*

the incomplete gamma function and its inverse, respectively. Further, let all  $\epsilon$ 's below be arbitrarily small constants.

1. Let  $\hat{\theta}_w$ , ( $\beta_w \leq \hat{\theta}_w \leq 1$ ) be the solution of

$$(1 - \epsilon_1^{(c)})(1 - \beta_w) \frac{\frac{\sqrt{2}\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \left(1 - \gamma_{inc}(\gamma_{inc}^{-1}(\frac{1-\theta_w}{1-\beta_w}, \frac{d}{2}), \frac{d+1}{2})\right)}{\theta_w} - \sqrt{2\gamma_{inc}^{-1}(\frac{(1+\epsilon_1^{(c)})(1-\theta_w)}{1-\beta_w}, \frac{d}{2})} = 0. \quad (6)$$

If  $\alpha$  and  $\beta_w$  further satisfy

$$\alpha d > (1 - \beta_w) \frac{2\Gamma(\frac{d+2}{2})}{\Gamma(\frac{d}{2})} \left(1 - \gamma_{inc}(\gamma_{inc}^{-1}(\frac{1-\hat{\theta}_w}{1-\beta_w}, \frac{d}{2}), \frac{d+2}{2})\right) + \beta_w d - \frac{\left((1 - \beta_w) \frac{\sqrt{2}\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} (1 - \gamma_{inc}(\gamma_{inc}^{-1}(\frac{1-\hat{\theta}_w}{1-\beta_w}, \frac{d}{2}), \frac{d+1}{2}))\right)^2}{\hat{\theta}_w} \quad (7)$$

then with overwhelming probability the solution of (5) is the  $k$ -block-sparse  $\mathbf{x}$  from (2).

2. Let  $\hat{\theta}_w$ , ( $\beta_w \leq \hat{\theta}_w \leq 1$ ) be the solution of

$$(1 + \epsilon_2^{(c)})(1 - \beta_w) \frac{\frac{\sqrt{2}\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \left(1 - \gamma_{inc}(\gamma_{inc}^{-1}(\frac{1-\theta_w}{1-\beta_w}, \frac{d}{2}), \frac{d+1}{2})\right)}{\theta_w} - \sqrt{2\gamma_{inc}^{-1}(\frac{(1-\epsilon_2^{(c)})(1-\theta_w)}{1-\beta_w}, \frac{d}{2})} = 0. \quad (8)$$

If  $\alpha$  and  $\beta_w$  further satisfy

$$\alpha d < \frac{1}{(1 + \epsilon_1^{(m)})^2} ((1 - \epsilon_1^{(g)})(1 - \beta_w) \frac{2\Gamma(\frac{d+2}{2})}{\Gamma(\frac{d}{2})} \left(1 - \gamma_{inc}(\gamma_{inc}^{-1}(\frac{1-\hat{\theta}_w}{1-\beta_w}, \frac{d}{2}), \frac{d+2}{2})\right) + \beta_w d - \frac{\left((1 - \beta_w) \frac{\sqrt{2}\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} (1 - \gamma_{inc}(\gamma_{inc}^{-1}(\frac{1-\hat{\theta}_w}{1-\beta_w}, \frac{d}{2}), \frac{d+1}{2}))\right)^2}{\hat{\theta}_w (1 + \epsilon_3^{(g)})^{-2}}) \quad (9)$$

then with overwhelming probability there will be a  $k$ -block-sparse  $\mathbf{x}$  (from a set of  $\mathbf{x}$ 's with fixed locations and directions of nonzero blocks) that satisfies (2) and is **not** the solution of (5).

*Proof.* The first part was established in [26]. The second part was established in [30].  $\square$

To give a little bit of flavor as to what is actually proven in the above theorem we in Figure 1 show the theoretical thresholds one obtains based on the above theorem. Moreover, in [26, 30] we showed that the theoretical prediction given in Figure 1 (and obtained in asymptotic infinite-dimensional scenario) are also in a solid agreement with what the finite dimensional numerical simulations can produce (in fact the agreement is already fairly good for the block-length as small as 100).

### 3 Large deviations

In the previous section we recalled on a collection of results that deal with the so-called phase-transition phenomenon. In this section we will introduce a somewhat novel concept that effectively resembles what is in probability theory known as the large deviation principle (LDP). We start by recalling on a couple of results that we established in [26].

For the simplicity and clarity of the exposition and without loss of generality we will assume that the blocks  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{n-k}$  of  $\mathbf{x}$  are equal to zero and that that vectors  $\mathbf{X}_{n-k+1}, \mathbf{X}_{n-k+2}, \dots, \mathbf{X}_n$  have fixed directions. Furthermore, since all probability distributions of interest will be rotationally invariant we will later assume that  $\mathbf{X}_i = (\|\mathbf{X}_i\|_2, 0, 0, \dots, 0)$ ,  $n - k + 1 \leq i \leq n$ . The following was proved in [26].

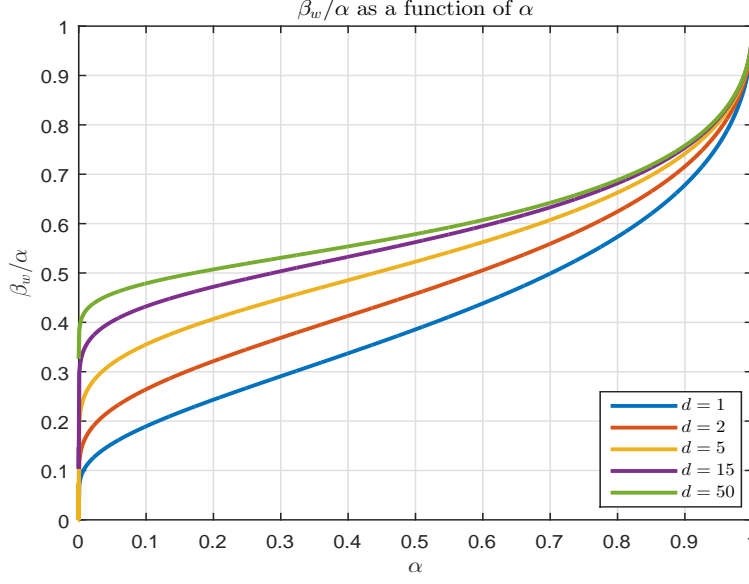


Figure 1: Block-sparse weak PT curves

**Theorem 2.** ([26] Nonzero blocks of  $\mathbf{x}$  have fixed directions and location) Assume that an  $M \times dn$  measurement matrix  $A$  is given. Let  $\mathbf{x}$  be a  $k$ -block-sparse vector. Also let  $\mathbf{X}_1 = \mathbf{X}_2 = \dots = \mathbf{X}_{n-k} = 0$ . Let the directions of vectors  $\mathbf{X}_{n-k+1}, \mathbf{X}_{n-k+2}, \dots, \mathbf{X}_n$  be fixed. Further, assume that  $\mathbf{y} = A\mathbf{x}$  and that  $\mathbf{w}$  is a  $dn \times 1$  vector. Then (5) will produce the solution of (2) if

$$(\forall \mathbf{w} \in \mathbf{R}^{dn} | A\mathbf{w} = 0) \quad - \quad \sum_{i=n-k+1}^n \frac{\mathbf{X}_i^T \mathbf{w}_i}{\|\mathbf{X}_i\|_2} < \sum_{i=1}^{n-k} \|\mathbf{w}_i\|_2. \quad (10)$$

To facilitate the exposition we set

$$S'_w = \{\mathbf{w} \in S^{dn-1} | \quad - \quad \sum_{i=n-k+1}^n \frac{\mathbf{X}_i^T \mathbf{w}_i}{\|\mathbf{X}_i\|_2} < \sum_{i=1}^{n-k} \|\mathbf{w}_i\|_2\}. \quad (11)$$

### 3.1 Upper tail

Assuming as earlier that the elements of  $A$  are i.i.d. standard normals, our object of interest in this section will be the following probability

$$P_{err} = P(\min_{\mathbf{w} \in S'_w} \|A\mathbf{w}\|_2 \leq 0) = P(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} (\mathbf{y}^T A\mathbf{w}) \geq 0). \quad (12)$$

Clearly,  $P_{err}$  is the probability that (5) fails to produce the solution of (2). Using the Chernoff bound we then get

$$P_{err} = P(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} (\mathbf{y}^T A\mathbf{w}) \geq 0) \leq E e^{c_3 \max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} (\mathbf{y}^T A\mathbf{w})} = E \max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{(-c_3 \mathbf{y}^T A\mathbf{w})}, \quad (13)$$

where we assume  $c_3 \geq 0$  and note that  $A$  and  $-A$  have the same distribution. We will then rely on the following lemma utilized in [28] (as discussed in [28], the lemma follows from a comparison result from [19]).

**Lemma 1.** ([19, 28]) Let  $A$  be an  $M \times N$  matrix with i.i.d. standard normal components. Let  $\mathbf{g}$  and  $\mathbf{h}$  be  $M \times 1$  and  $N \times 1$  vectors, respectively, with i.i.d. standard normal components. Also, let  $g$  be a standard

normal random variable and let  $c_3$  be a positive constant. Then

$$E\left(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{-c_3(\mathbf{y}^T \mathbf{A}\mathbf{w}+g)}\right) \leq E\left(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{-c_3(\mathbf{g}^T \mathbf{y}+\mathbf{h}^T \mathbf{w})}\right). \quad (14)$$

Transforming (12) a bit we obtain

$$P_{err} \leq E \max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{-c_3 \mathbf{y}^T \mathbf{A}\mathbf{w}} = e^{-\frac{c_3^2}{2}} E e^{c_3 g} E \max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{-c_3 \mathbf{y}^T \mathbf{A}\mathbf{w}} = e^{-\frac{c_3^2}{2}} E \max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{-c_3(\mathbf{y}^T \mathbf{A}\mathbf{w}+g)}. \quad (15)$$

Now a combination of (14) and (15) gives

$$\begin{aligned} P_{err} &\leq e^{-\frac{c_3^2}{2}} E \max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{-c_3(\mathbf{y}^T \mathbf{A}\mathbf{w}+g)} \leq e^{-\frac{c_3^2}{2}} E \max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} e^{-c_3(\mathbf{g}^T \mathbf{y}+\mathbf{h}^T \mathbf{w})} \\ &= e^{-\frac{c_3^2}{2}} E e^{-c_3 \|\mathbf{g}\|_2} E \max_{\mathbf{w} \in S'_w} e^{-c_3 \mathbf{h}^T \mathbf{w}} = e^{-\frac{c_3^2}{2}} E e^{-c_3 \|\mathbf{g}\|_2} E e^{c_3 \max_{\mathbf{w} \in S'_w} \mathbf{h}^T \mathbf{w}} = e^{-\frac{c_3^2}{2}} E e^{-c_3 \|\mathbf{g}\|_2} E e^{c_3 w(\mathbf{h}, S'_w)}, \end{aligned} \quad (16)$$

where

$$w(\mathbf{h}, S'_w) = \max_{\mathbf{w} \in S'_w} (\mathbf{h}^T \mathbf{w}) \quad (17)$$

and we also note that  $\mathbf{h}$  and  $-\mathbf{h}$  have the same distribution. We now set

$$S_w = \{\mathbf{w} \in S^{dn-1} \mid -\sum_{i=n-k+1}^n \mathbf{w}_{(i-1)d+1} < \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2\} \quad (18)$$

and

$$w(\mathbf{h}, S_w) = \max_{\mathbf{w} \in S_w} (\mathbf{h}^T \mathbf{w}). \quad (19)$$

Similarly to what was argued in [26] we have

$$P_{err} \leq e^{-\frac{c_3^2}{2}} E e^{-c_3 \|\mathbf{g}\|_2} E e^{c_3 w(\mathbf{h}, S'_w)} = e^{-\frac{c_3^2}{2}} E e^{-c_3 \|\mathbf{g}\|_2} E e^{c_3 w(\mathbf{h}, S_w)}. \quad (20)$$

For a moment we will now step away from  $P_{err}$  and will instead focus on  $w(\mathbf{h}, S_w)$ . As in [26] we have

$$w(\mathbf{h}, S_w) = \max_{\mathbf{w} \in S_w} (\mathbf{h}^T \mathbf{w}) = \max_{\mathbf{w} \in S_w} \left( \sum_{i=n-k+1}^n \mathbf{h}_{(i-1)d+1} \mathbf{w}_{(i-1)d+1} + \sum_{i=n-k+1}^n \|\mathbf{H}_i^*\|_2 \|\mathbf{W}_i^*\|_2 + \sum_{i=1}^{n-k} \|\mathbf{H}_i\|_2 \|\mathbf{W}_i\|_2 \right), \quad (21)$$

where

$$\begin{aligned} \mathbf{H}_i &= (\mathbf{h}_{(i-1)d+1}, \mathbf{h}_{(i-1)d+2}, \dots, \mathbf{h}_{id})^T, i = 1, 2, \dots, n-k \\ \mathbf{H}_i^* &= (\mathbf{h}_{(i-1)d+2}, \mathbf{h}_{(i-1)d+3}, \dots, \mathbf{h}_{id})^T, i = n-k+1, n-k+2, \dots, n \\ \mathbf{W}_i^* &= (\mathbf{w}_{(i-1)d+2}, \mathbf{w}_{(i-1)d+3}, \dots, \mathbf{w}_{id})^T, i = n-k+1, n-k+2, \dots, n. \end{aligned} \quad (22)$$

Set

$$\bar{\mathbf{H}} = (\|\mathbf{H}_1\|_2, \|\mathbf{H}_2\|_2, \dots, \|\mathbf{H}_{n-k}\|_2, -\mathbf{h}_{(n-k)d+1}, -\mathbf{h}_{(n-k+1)d+1}, -\mathbf{h}_{(n-k+2)d+1}, \dots, -\mathbf{h}_{(n-1)d+1}, \|\mathbf{H}_{n-k+1}^*\|_2, \|\mathbf{H}_{n-k+2}^*\|_2, \dots, \|\mathbf{H}_n^*\|_2)^T. \quad (23)$$

Then one can simplify (21) in the following way

$$w(\mathbf{h}, S_w) = \max_{\bar{\mathbf{y}} \in R^{n+k}} \sum_{i=1}^{n+k} \bar{\mathbf{H}}_i \bar{\mathbf{y}}_i$$

$$\begin{aligned}
\text{subject to} \quad & \bar{\mathbf{y}}_i \geq 0, 0 \leq i \leq n-k, n+1 \leq i \leq n+k \\
& \sum_{i=n-k+1}^n \bar{\mathbf{y}}_i \geq \sum_{i=1}^{n-k} \bar{\mathbf{y}}_i \\
& \sum_{i=1}^{n+k} \bar{\mathbf{y}}_i^2 \leq 1
\end{aligned} \tag{24}$$

where  $\bar{\mathbf{H}}_i$  is the  $i$ -th element of  $\bar{\mathbf{H}}$  and  $\bar{\mathbf{y}}_i$  is the  $i$ -th element of  $\bar{\mathbf{y}}$ . Utilizing the machinery of [26, 32] one then has

$$\begin{aligned}
w(\mathbf{h}, S_w) = & - \max_{\lambda \geq 0, \gamma \geq 0} \min_{\bar{\mathbf{y}}} \sum_{i=1}^{n+k} -\bar{\mathbf{H}}_i \bar{\mathbf{y}}_i + \lambda \sum_{i=1}^{n-k} \bar{\mathbf{y}}_i - \lambda \sum_{i=n-k+1}^n \bar{\mathbf{y}}_i + \gamma \sum_{i=1}^{n+k} \bar{\mathbf{y}}_i^2 - \gamma \\
\text{subject to} \quad & \bar{\mathbf{y}}_i \geq 0, 0 \leq i \leq n-k, n+1 \leq i \leq n+k.
\end{aligned} \tag{25}$$

After solving the inner minimization one finally obtains

$$\begin{aligned}
w(\mathbf{h}, S_w) &= \min_{\lambda \geq 0, \gamma \geq 0} \frac{\sum_{i=1}^{n-k} \max(\bar{\mathbf{H}}_i - \lambda, 0)^2 + \sum_{i=n-k+1}^n (\bar{\mathbf{H}}_i + \lambda)^2 + \sum_{i=n+1}^{n+k} \bar{\mathbf{H}}_i^2}{4\gamma} + \gamma \\
&= \min_{\lambda \geq 0} \sqrt{\sum_{i=1}^{n-k} \max(\bar{\mathbf{H}}_i - \lambda, 0)^2 + \sum_{i=n-k+1}^n (\bar{\mathbf{H}}_i + \lambda)^2 + \sum_{i=n+1}^{n+k} \bar{\mathbf{H}}_i^2}.
\end{aligned} \tag{26}$$

A combination of (20) and (26) provides an upper bound on  $P_{err}$ . We summarize these results in the following theorem.

**Theorem 3.** *Let  $A$  be an  $M \times N$  matrix in (2) with i.i.d. standard normal components. Let the unknown  $\mathbf{x}$  in (2) be  $k$ -block-sparse with the length of its blocks  $d$ . Further, let the location and the directions of nonzero blocks of  $\mathbf{x}$  be arbitrarily chosen but fixed. Let  $P_{err}$  be the probability that the solution of (5) is not the  $k$ -block-sparse solution of (2). Then*

$$P_{err} \leq \min_{c_3 \geq 0} e^{-\frac{c_3^2}{2}} e^{-c_3 \|\mathbf{g}\|^2} E e^{c_3 w(\mathbf{h}, S'_w)} = \min_{c_3 \geq 0} \left( e^{-\frac{c_3^2}{2}} \frac{1}{\sqrt{2\pi}^M} \int_{\mathbf{g}} e^{-\sum_{i=1}^M \mathbf{g}_i^2 / 2 - c_3 \|\mathbf{g}\|^2} d\mathbf{g} \min_{\lambda \geq 0, \gamma \geq \frac{c_3}{2}} w_1^{n-k} w_2^k w_3^k e^{c_3 \gamma} \right), \tag{27}$$

where

$$\begin{aligned}
w_1 &= \int_{\bar{\mathbf{H}}_i \geq 0} \frac{2^{-d/2}}{\Gamma(d/2)} \bar{\mathbf{H}}_i^{d/2-1} e^{-\bar{\mathbf{H}}_i/2} e^{c_3 \max(\sqrt{\bar{\mathbf{H}}_i - \lambda}, 0)^2 / 4 / \gamma} d\bar{\mathbf{H}}_i \\
w_2 &= \frac{1}{\sqrt{2\pi}} \int_{\bar{\mathbf{H}}_i} e^{-\bar{\mathbf{H}}_i^2/2} e^{c_3 (\bar{\mathbf{H}}_i + \lambda)^2 / 4 / \gamma} d\bar{\mathbf{H}}_i \\
w_3 &= \int_{\bar{\mathbf{H}}_i \geq 0} \frac{2^{-(d-1)/2}}{\Gamma((d-1)/2)} \bar{\mathbf{H}}_i^{(d-1)/2-1} e^{-\bar{\mathbf{H}}_i/2} e^{c_3 \bar{\mathbf{H}}_i / 4 / \gamma} d\bar{\mathbf{H}}_i.
\end{aligned} \tag{28}$$

*Proof.* Follows by combining (20) and (26) and noting that  $\bar{\mathbf{H}}_i, i \in [n-k+1, n]$ , are i.i.d. standard normals and  $\bar{\mathbf{H}}_i^2, i \notin [n-k+1, n]$ , are  $\chi$ -square distributed according to

$$\begin{aligned}
p(\bar{\mathbf{H}}_i^2) &= \frac{2^{-d/2}}{\Gamma(d/2)} (\bar{\mathbf{H}}_i^2)^{d/2-1} e^{-\bar{\mathbf{H}}_i^2/2}, 1 \leq i \leq n-k \\
p(\bar{\mathbf{H}}_i) &= \frac{1}{\sqrt{2\pi}} e^{-\bar{\mathbf{H}}_i^2/2}, n-k+1 \leq i \leq n \\
p(\bar{\mathbf{H}}_i^2) &= \frac{2^{-(d-1)/2}}{\Gamma((d-1)/2)} (\bar{\mathbf{H}}_i^2)^{(d-1)/2-1} e^{-\bar{\mathbf{H}}_i^2/2}, n+1 \leq i \leq n+k,
\end{aligned} \tag{29}$$

where, as usual,  $\Gamma(\cdot)$  stands for the standard gamma function.  $\square$

We emphasize here that such a bound is valid for any integers  $M$ ,  $k$ ,  $n$ , and  $d$  (of course assuming  $dk \leq M \leq dn$  so that the results make sense). As in the previous section, in this section we are also interested in the asymptotic scenario, namely the scenario from Theorem 2. In particular, we are interested in the scenarios where  $\alpha = \frac{m}{n} \geq (1 + \epsilon_\alpha)\alpha_w$  with  $\epsilon_\alpha > 0$  and  $\alpha_w$  being the minimal  $\alpha$  such that for a fixed  $\beta_w$  the equations of Theorem 2 are satisfied. In such a scenario what primarily determines  $P_{err}$  is its exponent. Hence, one has

$$I_{err} = \lim_{n \rightarrow \infty} \frac{\log P_{err}}{n}. \quad (30)$$

This of course clearly resembles the so-called large deviation principle (LDP) with  $I_{err}$  emulating the so-called LDP's rate (indicator) function. Based on Theorem 3 we have the following LDP type of theorem.

**Theorem 4.** *Assume the setup of Theorem 3. Further, let integers  $M$ ,  $k$ , and  $n$  be large ( $dk \leq M \leq dn$ ). Also, assume the following scaling:  $c_3 = c_3^{(s)}\sqrt{n}$  and  $\gamma = \gamma^{(s)}\sqrt{n}$ . Then*

$$I_{err} = \lim_{n \rightarrow \infty} \frac{\log P_{err}}{n} \leq \min_{c_3^{(s)} \geq 0} \left( -\frac{(c_3^{(s)})^2}{2} + I_{sph} + \min_{\lambda \geq 0, \gamma^{(s)} \geq 0} ((1 - \beta_w) \log w_1 + \beta_w \log w_2 + \beta_w \log w_3 + c_3^{(s)} \gamma^{(s)}) \right), \quad (31)$$

where

$$\begin{aligned} I_{sph} &= \widehat{\gamma^{(s)}} c_3^{(s)} - \frac{\alpha d}{2} \log \left( 1 - \frac{c_3^{(s)}}{2\widehat{\gamma^{(s)}}} \right) \\ \widehat{\gamma^{(s)}} &= \frac{2c_3^{(s)} - \sqrt{4(c_3^{(s)})^2 + 16\alpha d}}{8} \\ w_1 &= \int_{\bar{\mathbf{H}}_i \geq 0} \frac{2^{-d/2}}{\Gamma(d/2)} \bar{\mathbf{H}}_i^{d/2-1} e^{-\bar{\mathbf{H}}_i/2} e^{c_3^{(s)} \max(\sqrt{\bar{\mathbf{H}}_i} - \lambda, 0)^2/4/\gamma^{(s)}} d\bar{\mathbf{H}}_i \\ w_2 &= \frac{1}{\sqrt{2\pi}} \int_{\bar{\mathbf{H}}_i} e^{-\bar{\mathbf{H}}_i^2/2} e^{c_3^{(s)} (\bar{\mathbf{H}}_i + \lambda)^2/4/\gamma^{(s)}} d\bar{\mathbf{H}}_i = e^{(c_3^{(s)}/4/\gamma^{(s)})\lambda^2/(1-2c_3^{(s)}/4/\gamma^{(s)})} / \sqrt{1 - 2c_3^{(s)}/4/\gamma^{(s)}} \\ w_3 &= \int_{\bar{\mathbf{H}}_i \geq 0} \frac{2^{-(d-1)/2}}{\Gamma((d-1)/2)} \bar{\mathbf{H}}_i^{(d-1)/2-1} e^{-\bar{\mathbf{H}}_i/2} e^{c_3^{(s)} \bar{\mathbf{H}}_i/4/\gamma^{(s)}} d\bar{\mathbf{H}}_i = \left( 1/\sqrt{1 - 2c_3^{(s)}/4/\gamma^{(s)}} \right)^{d-1}. \end{aligned} \quad (32)$$

*Proof.* Follows from Theorem 3 and by noting that in [28] we established

$$I_{sph} = \lim_{n \rightarrow \infty} \frac{1}{n} \log(E e^{-c_3^{(s)} \sqrt{n} \|\mathbf{g}\|^2}) = \widehat{\gamma^{(s)}} c_3^{(s)} - \frac{\alpha d}{2} \log \left( 1 - \frac{c_3^{(s)}}{2\widehat{\gamma^{(s)}}} \right), \quad (33)$$

where

$$\widehat{\gamma^{(s)}} = \frac{2c_3^{(s)} - \sqrt{4(c_3^{(s)})^2 + 16\alpha d}}{8}. \quad (34)$$

$\square$

In Figure 2 we present the estimates of the  $P_{err}$  exponents (i.e. of  $I_{err}$  – the LDP's rate function values) that can be obtained using Theorem 4. Additionally, in Tables 1 and 2 we present some of the values we used for  $c_3^{(s)}$ ,  $\gamma^{(s)}$ , and  $\lambda$  to compute  $I_{err,u}^{(ub)}$  as the estimate. Plots in Figure 2 indicate that increasing block-length may shorten the so-called transition zone. A similar conclusion will also appear as reasonable in finite dimension scenarios that we will discuss in the following section.



Table 1: A collection of values for  $c_3^{(s)}$ ,  $\gamma^{(s)}$ ,  $\lambda$ , and  $I_{err,u}^{(ub)}$  in Theorem 4;  $d = 2$ ,  $\beta = \frac{1}{3}$

$\alpha$	0.6500	0.7100	0.7700	0.8300	0.8900
$c_3^{(s)}$	0.1263	0.7007	1.3534	2.2095	3.4521
$\gamma^{(s)}$	0.6026	0.7963	1.0451	1.4010	1.9539
$\lambda$	0.8973	0.7932	0.6965	0.5883	0.4701
$I_{err,u}^{(ub)}$	<b>-0.0007</b>	<b>-0.0215</b>	<b>-0.0701</b>	<b>-0.1475</b>	<b>-0.2565</b>

Table 2: A collection of values for  $c_3^{(s)}$ ,  $\gamma^{(s)}$ ,  $\lambda$ , and  $I_{err,u}^{(ub)}$  in Theorem 4;  $d = 10$ ,  $\beta = \frac{1}{3}$

$\alpha$	0.5900	0.6400	0.6900	0.7400	0.7900
$c_3^{(s)}$	0.2334	1.1136	2.0377	3.0881	4.3441
$\gamma^{(s)}$	1.2743	1.5736	1.9182	2.3361	2.8622
$\lambda$	2.0307	1.8998	1.7647	1.6125	1.4488
$I_{err,u}^{(ub)}$	<b>-0.0032</b>	<b>-0.0702</b>	<b>-0.2199</b>	<b>-0.4498</b>	<b>-0.7622</b>

### 3.2 Lower tail

In this subsection we focus on the lower bound of the lower tail type of large deviations. The results that we will present below will essentially complement the results from the previous subsection. As earlier, let the elements of  $A$  be i.i.d. standard normals. We will below consider the following probability

$$\begin{aligned}
P_{cor} &= P\left(\min_{A\mathbf{w}=0, \|\mathbf{w}\|_2 \leq 1} \sum_{i=n-k+1}^n \frac{\mathbf{X}_i^T \mathbf{W}_i}{\|\mathbf{X}_i\|_2} + \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2 \geq 0\right) \\
&= P\left(\min_{\|\mathbf{w}\|_2 \leq 1} \max_{\nu} \sum_{i=n-k+1}^n \frac{\mathbf{X}_i^T \mathbf{W}_i}{\|\mathbf{X}_i\|_2} + \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2 + \nu^T A\mathbf{w} \geq 0\right) \\
&= P\left(\max_{\nu} \min_{\|\mathbf{w}\|_2 \leq 1} \sum_{i=n-k+1}^n \frac{\mathbf{X}_i^T \mathbf{W}_i}{\|\mathbf{X}_i\|_2} + \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2 + \nu^T A\mathbf{w} \geq 0\right) \\
&= P\left(\max_{\nu} \min_{\|\mathbf{w}\|_2=1} \sum_{i=n-k+1}^n \frac{\mathbf{X}_i^T \mathbf{W}_i}{\|\mathbf{X}_i\|_2} + \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2 + \nu^T A\mathbf{w} \geq 0\right) \\
&\leq P\left(\max_{\nu} \min_{\|\mathbf{w}\|_2=1} \sum_{i=n-k+1}^n \frac{\mathbf{X}_i^T \mathbf{W}_i}{\|\mathbf{X}_i\|_2} + \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2 + \nu^T A\mathbf{w} + (g - t_1)\|\nu\|_2 \geq 0\right) / P(g \geq t_1),
\end{aligned} \tag{35}$$

where  $P_{cor} = 1 - P_{err}$  is the probability that (5) does produce the solution of (2),  $t_1$  is a scalar that will be discussed later, and  $g$  is a standard normal random variable independent of  $A$  (it is probably clear, but for the completeness, we recall that in order to lower bound the lower tail of  $P_{err}$ , it is enough to upper-bound the complementary quantity  $P_{cor}$ ). We will then make use of the following lemma from [19].

**Lemma 2.** [19] *Let  $A$  be an  $M \times N$  matrix with i.i.d. standard normal components. Let  $\mathbf{g}$  and  $\mathbf{h}$  be  $M \times 1$  and  $N \times 1$  vectors, respectively, with i.i.d. standard normal components. Also, let  $g$  be a standard normal random variable. Then*

$$P\left(\max_{\nu \in \mathbb{R}^n \setminus 0} \min_{\|\mathbf{w}\|_2=1} (\nu^T A\mathbf{w} + \|\nu\|_2 g - \zeta_{\mathbf{w}, \nu, t_1, \mathbf{X}}) \geq 0\right) \leq P\left(\max_{\nu \in \mathbb{R}^n \setminus 0} \min_{\|\mathbf{w}\|_2=1} (\mathbf{g}^T \nu + \|\nu\|_2 \mathbf{h}^T \mathbf{w} - \zeta_{\mathbf{w}, \nu, t_1, \mathbf{X}}) \geq 0\right). \tag{36}$$

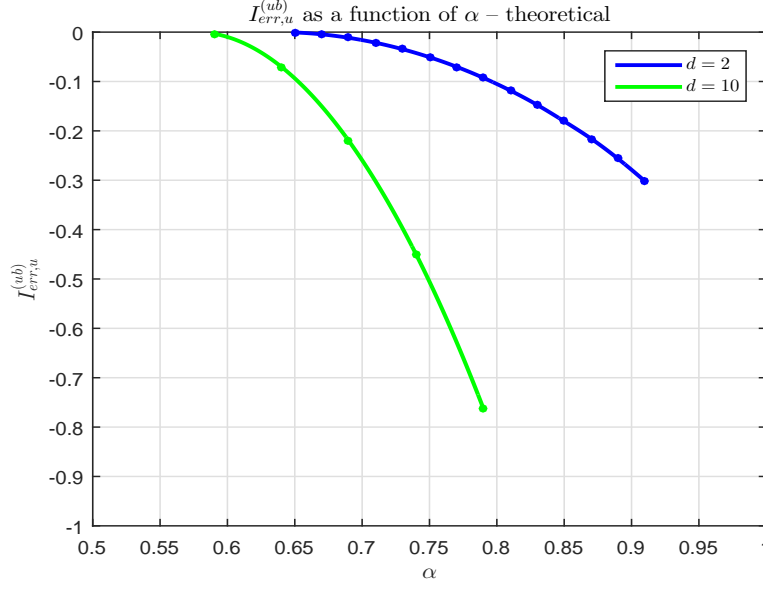


Figure 2:  $I_{err,u}^{(ub)}$  as a function of  $\alpha$ ; block-lengths  $d = 2$  and  $d = 10$

A combination of (35) and (36) (with a proper adjustment of  $\zeta_{\mathbf{w},\nu,t_1,\mathbf{x}}$ ) then gives

$$\begin{aligned}
P_{cor} &\leq P(\max_{\nu} \min_{\|\mathbf{w}\|_2=1} \sum_{i=n-k+1}^n \frac{\mathbf{x}_i^T \mathbf{W}_i}{\|\mathbf{x}_i\|_2} + \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2 + \nu^T A \mathbf{w} + (g - t_1) \|\nu\|_2 \geq 0) / P(g \geq t_1) \\
&\leq P(\max_{\lambda} \min_{\|\mathbf{w}\|_2=1} \lambda (\sum_{i=n-k+1}^n \frac{\mathbf{x}_i^T \mathbf{W}_i}{\|\mathbf{x}_i\|_2} + \sum_{i=1}^{n-k} \|\mathbf{W}_i\|_2) + \|\mathbf{g}\|_2 + \mathbf{h}^T \mathbf{w} - t_1 \geq 0) / P(g \geq t_1) \\
&\leq P(\|\mathbf{g}\|_2 - w(\mathbf{h}, S'_w) - t_1 \geq 0) / P(g \geq t_1) \\
&= P(\|\mathbf{g}\|_2 - w(\mathbf{h}, S_w) - t_1 \geq 0) / P(g \geq t_1),
\end{aligned} \tag{37}$$

where we set  $\lambda = 1/\|\nu\|_2$  and used the considerations from the previous subsection that relate to  $w(\mathbf{h}, S'_w)$  and  $w(\mathbf{h}, S_w)$ . Using the Chernoff bound we further have

$$P_{cor} \leq P(\|\mathbf{g}\|_2 - w(\mathbf{h}, S_w) - t_1 \geq 0) / P(g \geq t_1) \leq \min_{c_3 \geq 0} E e^{c_3 \|\mathbf{g}\|_2} E e^{-c_3 w(\mathbf{h}, S_w)} e^{-c_3 t_1} / P(g \geq t_1). \tag{38}$$

Using (38) one can then establish the following lower tail analogous version of Theorem 3 which provides an upper bound on  $P_{cor}$ .

**Theorem 5.** *Let  $A$  be an  $M \times N$  matrix in (2) with i.i.d. standard normal components. Let the unknown  $\mathbf{x}$  in (2) be  $k$ -block-sparse with the length of its blocks  $d$ . Further, let the location and the directions of nonzero blocks of  $\mathbf{x}$  be arbitrarily chosen but fixed. Let  $P_{cor}$  be the probability that the solution of (5) is the  $k$ -block-sparse solution of (2). Then*

$$P_{cor} \leq \min_{t_1} \min_{c_3 \geq 0} E e^{c_3 \|\mathbf{g}\|_2} E e^{-c_3 w(\mathbf{h}, S'_w)} e^{-c_3 t_1} / P(g \geq t_1). \tag{39}$$

We again emphasize that the above bound is valid for any integers  $M$ ,  $k$ ,  $n$ , and  $d$  (of course, as earlier, assuming  $dk \leq M \leq dn$  so that the results make sense). As in the previous section, in this section we are also interested in the asymptotic scenario, namely the scenario from Theorem 2. Differently from the previous

subsection, here we are interested in the scenarios where  $\alpha = \frac{m}{n} \geq (1 - \epsilon_\alpha)\alpha_w$  with  $\epsilon_\alpha > 0$  and  $\alpha_w$  being the minimal  $\alpha$  such that for a fixed  $\beta_w$  the equations of Theorem 2 are satisfied. Similarly to what we had earlier for the exponent of  $P_{err}$  we now have

$$I_{cor} = \lim_{n \rightarrow \infty} \frac{\log P_{cor}}{n}. \quad (40)$$

Based on Theorem 3 and similarly as in Theorem 4 we have the following LDP type of result.

**Theorem 6.** *Assume the setup of Theorem 5. Further, let integers  $M$ ,  $k$ , and  $n$  be large ( $dk \leq M \leq dn$ ). Also, assume the following scaling:  $c_3 = c_3^{(s)}\sqrt{n}$  and  $\gamma = \gamma^{(s)}\sqrt{n}$ . Then*

$$I_{cor} = \lim_{n \rightarrow \infty} \frac{\log P_{cor}}{n} \leq \min_{c_3^{(s)} \geq 0} \left( -\frac{(c_3^{(s)})^2}{2} + I_{sph}^+ + \max_{\lambda \geq 0, \gamma^{(s)} \geq 0} ((1 - \beta_w) \log w_1 + \beta_w \log w_2 + \beta_w \log w_3 - c_3^{(s)} \gamma^{(s)}) \right), \quad (41)$$

where

$$\begin{aligned} I_{sph}^+ &= \widehat{\gamma_+^{(s)}} c_3^{(s)} - \frac{\alpha d}{2} \log \left( 1 - \frac{c_3^{(s)}}{2\widehat{\gamma_+^{(s)}}} \right) \\ \widehat{\gamma_+^{(s)}} &= \frac{2c_3^{(s)} + \sqrt{4(c_3^{(s)})^2 + 16\alpha d}}{8} \\ w_1 &= \int_{\bar{\mathbf{H}}_i \geq 0} \frac{2^{-d/2}}{\Gamma(d/2)} \bar{\mathbf{H}}_i^{d/2-1} e^{-\bar{\mathbf{H}}_i/2} e^{-c_3^{(s)} \max(\sqrt{\bar{\mathbf{H}}_i} - \lambda, 0)^2/4/\gamma^{(s)}} d\bar{\mathbf{H}}_i \\ w_2 &= \frac{1}{\sqrt{2\pi}} \int_{\bar{\mathbf{H}}_i} e^{-\bar{\mathbf{H}}_i/2} e^{-c_3^{(s)} (\bar{\mathbf{H}}_i + \lambda)^2/4/\gamma^{(s)}} d\bar{\mathbf{H}}_i = e^{(-c_3^{(s)}/4/\gamma^{(s)})\lambda^2/(1+2c_3^{(s)}/4/\gamma^{(s)})} / \sqrt{1 + 2c_3^{(s)}/4/\gamma^{(s)}} \\ w_3 &= \int_{\bar{\mathbf{H}}_i \geq 0} \frac{2^{-d/2}}{\Gamma(d/2)} \bar{\mathbf{H}}_i^{d/2-1} e^{-\bar{\mathbf{H}}_i/2} e^{-c_3^{(s)} \bar{\mathbf{H}}_i/4/\gamma^{(s)}} d\bar{\mathbf{H}}_i = \left( 1/\sqrt{1 + 2c_3^{(s)}/4/\gamma^{(s)}} \right)^{d-1}. \end{aligned} \quad (42)$$

*Proof.* Follows from Theorem 5 optimizing over  $t_1$  and by noting that in [28] we established

$$I_{sph}^+ = \lim_{n \rightarrow \infty} \frac{1}{n} \log(E e^{c_3^{(s)} \sqrt{n} \|\mathbf{g}\|^2}) = \widehat{\gamma_+^{(s)}} c_3^{(s)} - \frac{\alpha d}{2} \log \left( 1 - \frac{c_3^{(s)}}{2\widehat{\gamma_+^{(s)}}} \right), \quad (43)$$

where

$$\widehat{\gamma_+^{(s)}} = \frac{2c_3^{(s)} + \sqrt{4(c_3^{(s)})^2 + 16\alpha d}}{8}, \quad (44)$$

and that analogously one has

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log(E e^{-c_3 w(\mathbf{h}, S'_w)}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left( E e^{-c_3 \min_{\lambda \geq 0, \gamma \geq 0} \left( \frac{\sum_{i=1}^{n-k} \max(\bar{\mathbf{H}}_i - \lambda, 0)^2 + \sum_{i=n-k+1}^n (\bar{\mathbf{H}}_i + \lambda)^2 + \sum_{i=n+1}^{n+k} \bar{\mathbf{H}}_i^2}{4\gamma} + \gamma \right)} \right) \\ &= \max_{\lambda \geq 0, \gamma^{(s)} \geq 0} ((1 - \beta_w) \log w_1 + \beta_w \log w_2 + \beta_w \log w_3 - c_3^{(s)} \gamma^{(s)}). \end{aligned} \quad (45)$$

□

In Figure 3 we present the estimates of the  $P_{cor}$  exponents (i.e. of  $I_{cor}$ ) that one can obtain using Theorem 6. Additionally, in Tables 3 and 4 we present some of the values we used for  $c_3^{(s)}$ ,  $\gamma^{(s)}$ , and  $\lambda$  to compute  $I_{cor,l}^{(ub)}$  as the estimate. As in the previous subsection, plots in Figure 3 indicate that increasing block-length may shorten the so-called transition zone.

Table 3: A collection of values for  $c_3^{(s)}$ ,  $\gamma^{(s)}$ ,  $\lambda$ , and  $I_{cor,l}^{(ub)}$  in Theorem 6;  $d = 2$ ,  $\beta = \frac{1}{3}$

$\alpha$	0.5700	0.5300	0.4900	0.4500	0.4100
$c_3^{(s)}$	0.6198	1.0547	1.5715	2.3026	3.5891
$\gamma^{(s)}$	0.4005	0.3128	0.2391	0.1703	0.1078
$\lambda$	1.0187	1.0810	1.1474	1.2214	1.3014
$I_{cor,l}^{(ub)}$	<b>-0.0185</b>	<b>-0.0495</b>	<b>-0.0978</b>	<b>-0.1674</b>	<b>-0.2649</b>

Table 4: A collection of values for  $c_3^{(s)}$ ,  $\gamma^{(s)}$ ,  $\lambda$ , and  $I_{cor,l}^{(ub)}$  in Theorem 6;  $d = 10$ ,  $\beta = \frac{1}{3}$

$\alpha$	0.5500	0.5100	0.4700	0.4300	0.3900
$c_3^{(s)}$	0.4789	1.2466	2.1555	3.4129	5.7283
$\gamma^{(s)}$	1.0590	0.8598	0.6717	0.4896	0.3075
$\lambda$	2.1338	2.2412	2.3487	2.4591	2.5873
$I_{cor,l}^{(ub)}$	<b>-0.0132</b>	<b>-0.0872</b>	<b>-0.2361</b>	<b>-0.4791</b>	<b>-0.8549</b>

## 4 Finite dimensions

In this section we will raise the level of considerations a bit (in fact quite a bit). Instead of basically neglecting the true systems dimensions (as was done in the previous sections) we will here specifically focus on them. When it comes to handling problems similar to those considered here, the theories that we have built over the years are fairly superior and we found hardly any other known mathematical tool that can outmatch them. The main features that essentially make them very hard to beat are their exactness, universality, and ultimately their simplicity. A bit lost among them is one additional feature that in a way distinguishes them. Namely, almost all of the results that we created are fairly easy to adjust to finite dimension scenarios. In such scenarios the original simplicity and universality/applicability are preserved. For the problem of interest in this paper, we will towards the end of this section demonstrate that this is indeed true. Before doing that we will focus on another path that we view as essentially the only counterpart that sometimes can match the performance characterization capabilities of the theories that we have built.

### 4.1 High-dimensional geometry

The path that we will consider below can be viewed as geometrical and as such is in a rapid contrast with the techniques that we employed in [26, 30–32] and ultimately in earlier sections of this paper. We will start things off by introducing a few terms and notions well known in high-dimensional geometry, again assuming a high degree of familiarity with some of them. To that end, let  $S_{st}$  be the intersection of a convex cone  $C_{st}$  and the  $n_{st}$ -dimensional unit sphere,  $S^{n_{st}-1}$ , i.e.

$$S_{st} = C_{st} \cap S^{n_{st}-1}. \quad (46)$$

Let  $\zeta_{\epsilon_{st}}(S_{st})$  be the  $\epsilon_{st}$ -neighborhood ( $0 < \epsilon_{st} < \pi/2$ ) spherical extension of  $S_{st}$ , i.e.

$$\zeta_{\epsilon_{st}}(S_{st}) = \{\mathbf{y} \in S^{n_{st}-1} \mid 0 < \min_{\mathbf{w} \in S_{st}} \arccos(\mathbf{y}^T \mathbf{w}) \leq \epsilon_{st}\}. \quad (47)$$

Then the following is the so-called spherical Steiner formula (see, e.g. [25])

$$\sigma_{n_{st}-1}(\zeta_{\epsilon_{st}}(S_{st})) = \sum_{i_{st}=0}^{n_{st}-2} g_{n_{st}, i_{st}} v_{i_{st}}(S_{st}), \quad (48)$$

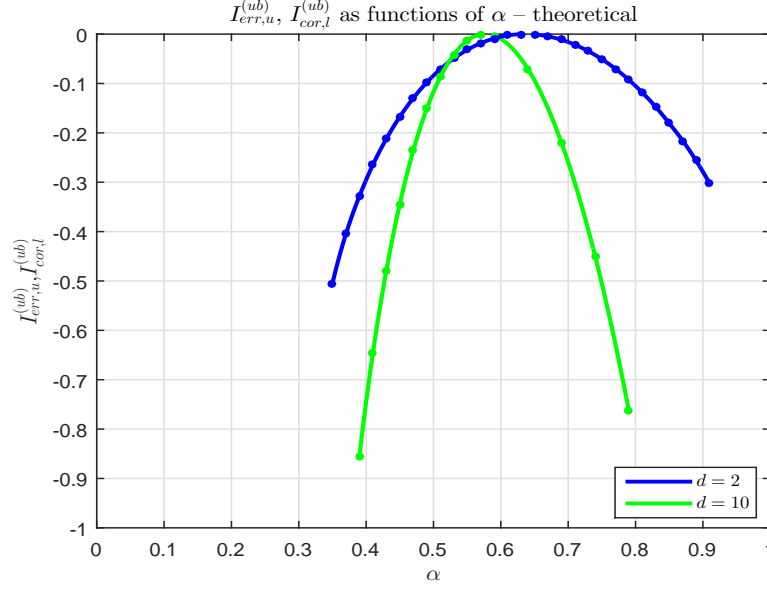


Figure 3:  $I_{err,u}^{(ub)}$ ,  $I_{cor,l}^{(ub)}$  as functions of  $\alpha$ ; block-lengths  $d = 2$  and  $d = 10$

where

$$g_{n_{st}, i_{st}} = \sigma_{i_{st}}(S^{i_{st}}) \sigma_{n_{st}-i_{st}-2}(S^{n_{st}-i_{st}-2}) \int_0^{\epsilon_{st}} \cos^{i_{st}} \phi \sin^{n_{st}-i_{st}-2} \phi d\phi, \quad (49)$$

and  $\sigma_{n_{st}-1}(\cdot)$  is the standard Lebesgue measure on the unit  $S^{n_{st}-1}$ . In fact,

$$\sigma_{n_{st}-1}(S^{n_{st}-1}) = \frac{2\pi^{n_{st}/2}}{\Gamma(\frac{n_{st}}{2})}. \quad (50)$$

The numbers  $v_{i_{st}}(S_{st})$ ,  $0 \leq i_{st} \leq n_{st} - 2$ , in the spherical Steiner formula are often supplemented by

$$v_{n_{st}-1}(S_{st}) = \frac{\sigma_{n_{st}-1}(S_{st})}{\sigma_{n_{st}-1}(S^{n_{st}-1})}. \quad (51)$$

One also assumes  $v_{i_{st}}(\emptyset) = 0$  and  $v_{-1}(S_{st}) = v_{n_{st}-1}(\mathcal{D}(S_{st}))$ , where

$$\mathcal{D}(S_{st}) = \{\mathbf{y} \in S^{n_{st}-1} \mid \mathbf{y}^T \mathbf{w} \leq 0 \quad \forall \mathbf{w} \in S_{st}\}. \quad (52)$$

The numbers  $v_{i_{st}}$  are uniquely determined by the Steiner spherical formula and are called the spherical intrinsic volumes. The following spherical Crofton like formula connects the spherical intrinsic volumes and the likelihoods that a random  $(n_{st}-j)$ -dimensional  $G_{n_{st}, n_{st}-j}$  linear subspace of  $R^{n_{st}}$  from the corresponding Grassmannian misses sets of  $S_{st}$  type (see, e.g. [25])

$$P(G_{n_{st}, n_{st}-j} \cap S_{st} \neq \emptyset) = 2 \sum_{k_{st}=0}^{\lfloor \frac{n_{st}-1-j}{2} \rfloor} v_{j+2k_{st}+1}(S_{st}). \quad (53)$$

Choosing  $S_{st} = S'_w$  and connecting (12) and (53) while keeping in mind a rotational invariance of  $A$  (alternatively one can instead of Gaussian  $A$ 's consider  $A$ 's that have the null-space distributed uniformly randomly

in the corresponding Grassmanians) one has

$$P_{err} = P(\min_{\mathbf{w} \in S'_w} \|\mathbf{A}\mathbf{w}\|_2 \leq 0) = P(G_{dn, dn-M} \cap S'_w \neq \emptyset) = 2 \sum_{k_{st}=0}^{\lfloor \frac{dn-1-M}{2} \rfloor} v_{M+2k_{st}+1}(S'_w). \quad (54)$$

The above formula then positions Crofton like (53) as a potentially powerful tool in determining  $P_{err}$ . In fact, the combination of (48) and (53) is indeed a well known principal tool in high-dimensional integral geometry. The problem is that it typically remains of not much practical use as it requires addressing several major problems. Resolving these problems is in general much harder than deriving any of (48) and (53) (we should though emphasize that while deriving (48) is basically trivial for high-dimensional integral geometry experts, deriving (53) is not super obvious and typically uses as key steps the ideas that date back to the works of Santalo, Hadwiger, and McMullen, obviously the integral geometry greats). Namely, to use (48) one should be able first to determine  $\sigma_{n_{st}-1}(\zeta_{\epsilon_{st}}(S_{st}))$  and then to hope that the system on the right side of (48) is (computationally speaking) well-posed. In the remaining parts of this section we will try to address these problems.

In general, determining  $\sigma_{n_{st}-1}(\zeta_{\epsilon_{st}}(S_{st}))$  analytically is rarely known to be possible. Below, we will present a strategy that heavily relies on the machinery that we introduced in [26, 27, 36, 37] and to a degree utilized in the previous sections of this paper. Using such a strategy we will be able to determine asymptotically optimal estimates of  $\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))$  that will also turn out to be quite close approximations in finite dimensions. Moreover, the approximations that we will derive will in fact be rigorous upper bounds on  $\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))$  (the numerical results that we will present will in fact confirm that these bounds are indeed very, very close to the exact values).

#### 4.1.1 Determining $\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$

We will find it a bit more natural to work with a scaled version of  $\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))$ , namely the ratio  $\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$  (of course, determining  $\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$  automatically determines  $\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))$ ). We start by quickly modifying the  $\zeta_{\epsilon_{st}}(S'_w)$  given in (47) with  $S_{st}$  obviously replaced by  $S'_w$ . To that end we have

$$\begin{aligned} \zeta_{\epsilon_{st}}(S'_w) &= \{\mathbf{y} \in S^{dn-1} \mid 0 < \min_{\mathbf{w} \in S'_w} \arccos(\mathbf{y}^T \mathbf{w}) \leq \epsilon_{st}\} \\ &= \{\mathbf{y} \in S^{dn-1} \mid 1 > \cos(\min_{\mathbf{w} \in S'_w} \arccos(\mathbf{y}^T \mathbf{w})) \geq \cos(\epsilon_{st})\} \\ &= \{\mathbf{y} \in S^{dn-1} \mid 1 > \max_{\mathbf{w} \in S'_w} \cos(\arccos(\mathbf{y}^T \mathbf{w})) \geq \cos(\epsilon_{st})\} \\ &= \{\mathbf{y} \in S^{dn-1} \mid \max_{\mathbf{w} \in S'_w} (\mathbf{y}^T \mathbf{w}) \geq \cos(\epsilon_{st})\}. \end{aligned} \quad (55)$$

We then impose the uniform measure over the unit sphere  $S^{dn-1}$  through the standard Gaussians. Namely, let, as earlier,  $\mathbf{h}$  be an  $dn$ -dimensional vector of i.i.d. standard normals. Then, due to the rotational invariance of  $\mathbf{h}$ , we have

$$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})} = \frac{\sigma_{dn-1}(\{\mathbf{y} \in S^{dn-1} \mid \max_{\mathbf{w} \in S'_w} (\mathbf{y}^T \mathbf{w}) \geq \cos(\epsilon_{st})\})}{\sigma_{dn-1}(S^{dn-1})} = P(\max_{\mathbf{w} \in S'_w} \frac{\mathbf{h}^T \mathbf{w}}{\|\mathbf{h}\|_2} \geq \cos(\epsilon_{st})). \quad (56)$$

Based on (17) and (56) we easily obtain

$$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})} = P((\max_{\mathbf{w} \in S'_w} (\mathbf{h}^T \mathbf{w}))^2 \geq \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2) = P((w(\mathbf{h}, S'_w))^2 \geq \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2). \quad (57)$$

Similarly to what was done in (20) and ultimately based on the rotation considerations from [26] we also have

$$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})} = P((w(\mathbf{h}, S'_w))^2 \geq \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2) = P((w(\mathbf{h}, S_w))^2 \geq \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2). \quad (58)$$

Now, set

$$\begin{aligned}
w_\lambda(\mathbf{h}, S_w) &= \sqrt{\sum_{i=1}^{n-k} \max(\bar{\mathbf{H}}_i - \lambda, 0)^2 + \sum_{i=n-k+1}^n (\bar{\mathbf{H}}_i + \lambda)^2 + \sum_{i=n+1}^{n+k} \bar{\mathbf{H}}_i^2} \\
w_{\lambda,1} &= \sum_{i=1}^{n-k} \max(\bar{\mathbf{H}}_i - \lambda, 0)^2 \\
w_{\lambda,2} &= \sum_{i=n-k+1}^n (\bar{\mathbf{H}}_i + \lambda)^2 \\
w_{\lambda,3} &= \sum_{i=n+1}^{n+k} \bar{\mathbf{H}}_i^2.
\end{aligned} \tag{59}$$

Then a combination of (58) and (26) further gives

$$\begin{aligned}
\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})} &= P((w(\mathbf{h}, S_w))^2 \geq \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2) \\
&= P((\min_{\lambda \geq 0} w_\lambda(\mathbf{h}, S_w))^2 \geq \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2) \\
&\leq \min_{\lambda \geq 0} P((w_\lambda(\mathbf{h}, S_w))^2 \geq \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2) \\
&= 1 - \max_{\lambda \geq 0} P((w_\lambda(\mathbf{h}, S_w))^2 < \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2).
\end{aligned} \tag{60}$$

We will now focus on computing  $P((w_\lambda(\mathbf{h}, S_w))^2 < \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2)$ . We start by observing the following simple identity

$$P_\lambda = P((w_\lambda(\mathbf{h}, S_w))^2 < \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2) = \frac{1}{\sqrt{2\pi}^{dn}} \int_{\mathbf{h}} h((-w_\lambda(\mathbf{h}, S_w))^2 + \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2) e^{-\frac{\|\mathbf{h}\|_2^2}{2}} d\mathbf{h}, \tag{61}$$

where  $h(t)$  is the standard indicator step function. The following characterization of the standard step function we found fairly useful

$$h(t) = \lim_{\epsilon_{wf} \rightarrow 0_+} \int_{-\infty}^{\infty} \frac{e^{jw_f t}}{2\pi j(w_f - j\epsilon_{wf})} dw_f. \tag{62}$$

Combining (61) and (62) we then have

$$\begin{aligned}
P_\lambda &= \frac{1}{\sqrt{2\pi}^{dn}} \int_{\mathbf{h}} h((-w_\lambda(\mathbf{h}, S_w))^2 + \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2 \geq 0) e^{-\frac{\|\mathbf{h}\|_2^2}{2}} d\mathbf{h} \\
&= \frac{1}{\sqrt{2\pi}^{dn}} \int_{\mathbf{h}} \lim_{\epsilon_{wf} \rightarrow 0_+} \int_{-\infty}^{\infty} \frac{e^{-jw_f((w_\lambda(\mathbf{h}, S_w))^2 - \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2)}}{2\pi j(w_f - j\epsilon_{wf})} dw_f e^{-\frac{\|\mathbf{h}\|_2^2}{2}} d\mathbf{h} \\
&= \lim_{\epsilon_{wf} \rightarrow 0_+} \int_{-\infty}^{\infty} \frac{1}{2\pi j(w_f - j\epsilon_{wf})} \frac{1}{\sqrt{2\pi}^{dn}} \int_{\mathbf{h}} e^{-jw_f(w_{\lambda,1} + w_{\lambda,2} + w_{\lambda,3} - \cos^2(\epsilon_{st}) \|\mathbf{h}\|_2^2)} e^{-\frac{\|\mathbf{h}\|_2^2}{2}} d\mathbf{h} dw_f \\
&= \lim_{\epsilon_{wf} \rightarrow 0_+} \int_{-\infty}^{\infty} \frac{1}{2\pi j(w_f - j\epsilon_{wf})} I_{\lambda,1}(w_f, \epsilon_{st}) I_{\lambda,2}(w_f, \epsilon_{st}) I_{\lambda,3}(w_f, \epsilon_{st}) dw_f,
\end{aligned} \tag{63}$$

where

$$I_{\lambda,1}(w_f, \epsilon_{st}) = \frac{1}{\sqrt{2\pi}^{d(n-k)}} \int e^{-jw_f(w_{\lambda,1} - \cos^2(\epsilon_{st}) \|\mathbf{h}_{1:d(n-k)}\|_2^2)} e^{-\frac{\|\mathbf{h}_{1:d(n-k)}\|_2^2}{2}} d\mathbf{h}_{1:d(n-k)}$$

$$\begin{aligned}
I_{\lambda,2}(w_f, \epsilon_{st}) &= \frac{1}{\sqrt{2\pi}^k} \int e^{-jw_f(w_{\lambda,2} - \cos^2(\epsilon_{st}) \sum_{i=n-k+1}^n \|\mathbf{h}_{d(i-1)+1}\|_2^2)} \prod_{i=n-k+1}^n e^{-\frac{\|\mathbf{h}_{d(i-1)+1}\|_2^2}{2}} d\mathbf{h}_{d(i-1)+1} \\
I_{\lambda,3}(w_f, \epsilon_{st}) &= \frac{1}{\sqrt{2\pi}^{(d-1)(k)}} \int e^{-jw_f(w_{\lambda,3} - \cos^2(\epsilon_{st}) \sum_{i=n-k+1}^n \|\mathbf{h}_{d(i-1)+2:di}\|_2^2)} \prod_{i=n-k+1}^n e^{-\frac{\|\mathbf{h}_{d(i-1)+2:di}\|_2^2}{2}} d\mathbf{h}_{d(i-1)+2:di}.
\end{aligned} \tag{64}$$

We now analyze separately each of  $I_{\lambda,1}(w_f, \epsilon_{st})$ ,  $I_{\lambda,2}(w_f, \epsilon_{st})$ , and  $I_{\lambda,3}(w_f, \epsilon_{st})$ .

1) **Computing**  $I_{\lambda,1}(w_f, \epsilon_{st})$

From (22) we recall that  $\bar{\mathbf{H}}_i = (\mathbf{h}_{(i-1)d+1}, \mathbf{h}_{(i-1)d+2}, \dots, \mathbf{h}_{id})^T$ ,  $i = 1, 2, \dots, n-k$ , and from (23) we recall that  $\bar{\mathbf{H}}_i = \|\mathbf{H}_i\|_2$ ,  $i = 1, 2, \dots, n-k$ . Then we have

$$\begin{aligned}
I_{\lambda,1}(w_f, \epsilon_{st}) &= \frac{1}{\sqrt{2\pi}^{d(n-k)}} \int e^{-jw_f(w_{\lambda,1} - \cos^2(\epsilon_{st}) \|\mathbf{h}_{1:d(n-k)}\|_2^2)} e^{-\frac{\|\mathbf{h}_{1:d(n-k)}\|_2^2}{2}} d\mathbf{h}_{1:d(n-k)} \\
&= \frac{1}{\sqrt{2\pi}^{d(n-k)}} \int e^{-jw_f(\sum_{i=1}^{n-k} \max(\bar{\mathbf{H}}_i - \lambda, 0)^2 - \cos^2(\epsilon_{st}) \|\mathbf{h}_{1:d(n-k)}\|_2^2)} e^{-\frac{\|\mathbf{h}_{1:d(n-k)}\|_2^2}{2}} d\mathbf{h}_{1:d(n-k)} \\
&= \frac{1}{\sqrt{2\pi}^{d(n-k)}} \int e^{-jw_f(\sum_{i=1}^{n-k} \max(\bar{\mathbf{H}}_i - \lambda, 0)^2 - \cos^2(\epsilon_{st}) \sum_{i=1}^{n-k} \bar{\mathbf{H}}_i^2)} e^{-\frac{\sum_{i=1}^{n-k} \bar{\mathbf{H}}_i^2}{2}} d\mathbf{h}_{1:d(n-k)} \\
&= \frac{1}{\sqrt{2\pi}^{d(n-k)}} \int \prod_{i=1}^{n-k} e^{-jw_f(\max(\bar{\mathbf{H}}_i - \lambda, 0)^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_i^2)} e^{-\frac{\bar{\mathbf{H}}_i^2}{2}} d\mathbf{h}_{(i-1)d+1:id} \\
&= \prod_{i=1}^{n-k} \frac{1}{\sqrt{2\pi}^d} \int e^{-jw_f(\max(\bar{\mathbf{H}}_i - \lambda, 0)^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_i^2)} e^{-\frac{\bar{\mathbf{H}}_i^2}{2}} d\mathbf{h}_{(i-1)d+1:id} \\
&= \left( \frac{1}{\sqrt{2\pi}^d} \int e^{-jw_f(\max(\bar{\mathbf{H}}_i - \lambda, 0)^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_i^2)} e^{-\frac{\bar{\mathbf{H}}_i^2}{2}} d\mathbf{h}_{(i-1)d+1:id} \right)^{n-k}, 1 \leq i \leq n-k. \tag{65}
\end{aligned}$$

Using the distribution function for  $\bar{\mathbf{H}}_i$ ,  $1 \leq i \leq n-k$ , from (29) we further have

$$\begin{aligned}
I_{\lambda,1}(w_f, \epsilon_{st}) &= \left( \frac{1}{\sqrt{2\pi}^d} \int e^{-jw_f(\max(\bar{\mathbf{H}}_i - \lambda, 0)^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_i^2)} e^{-\frac{\bar{\mathbf{H}}_i^2}{2}} d\mathbf{h}_{(i-1)d+1:id} \right)^{n-k} \\
&= \left( \int_{\bar{\mathbf{H}}_i \geq 0} e^{-jw_f(\max(\sqrt{\bar{\mathbf{H}}_i} - \lambda, 0)^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_i)} \frac{2^{-d/2}}{\Gamma(d/2)} \bar{\mathbf{H}}_i^{d/2-1} e^{-\bar{\mathbf{H}}_i/2} d\bar{\mathbf{H}}_i \right)^{n-k}. \tag{66}
\end{aligned}$$

2) **Computing**  $I_{\lambda,2}(w_f, \epsilon_{st})$

From (23) we recall that  $\bar{\mathbf{H}}_{n-k+1:n} = (-\mathbf{h}_{(n-k)d+1}, -\mathbf{h}_{(n-k+1)d+1}, \dots, -\mathbf{h}_{(n-1)d+1})$ , i.e.  $\bar{\mathbf{H}}_i = -\mathbf{h}_{(i-1)d+1}$ ,  $i = n-k+1, n-k+2, \dots, n$ . Then we have

$$\begin{aligned}
I_{\lambda,2}(w_f, \epsilon_{st}) &= \frac{1}{\sqrt{2\pi}^k} \int e^{-jw_f(w_{\lambda,2} - \cos^2(\epsilon_{st}) \sum_{i=n-k+1}^n \mathbf{h}_{d(i-1)+1}^2)} \prod_{i=n-k+1}^n e^{-\frac{\mathbf{h}_{d(i-1)+1}^2}{2}} d\mathbf{h}_{d(i-1)+1} \\
&= \frac{1}{\sqrt{2\pi}^k} \int e^{-jw_f(\sum_{i=n-k+1}^n (\mathbf{H}_i + \lambda)^2 - \cos^2(\epsilon_{st}) \sum_{i=n-k+1}^n \mathbf{h}_{d(i-1)+1}^2)} \prod_{i=n-k+1}^n e^{-\frac{\mathbf{h}_{d(i-1)+1}^2}{2}} d\mathbf{h}_{d(i-1)+1} \\
&= \frac{1}{\sqrt{2\pi}^k} \int \prod_{i=n-k+1}^n e^{-jw_f((- \mathbf{h}_{d(i-1)+1} + \lambda)^2 - \cos^2(\epsilon_{st}) \mathbf{h}_{d(i-1)+1}^2)} e^{-\frac{\mathbf{h}_{d(i-1)+1}^2}{2}} d\mathbf{h}_{d(i-1)+1} \\
&= \left( \frac{1}{\sqrt{2\pi}} \int e^{-jw_f((- \mathbf{h}_{d(i-1)+1} + \lambda)^2 - \cos^2(\epsilon_{st}) \mathbf{h}_{d(i-1)+1}^2)} e^{-\frac{\mathbf{h}_{d(i-1)+1}^2}{2}} d\mathbf{h}_{d(i-1)+1} \right)^k
\end{aligned}$$



$$= \left( \exp \left( -jw_f \lambda^2 + \left( 2\lambda jw_f / \sqrt{1 - 2jw_f (\cos^2(\epsilon_{st}) - 1)} \right)^2 / 2 \right) / \sqrt{1 - 2jw_f (\cos^2(\epsilon_{st}) - 1)} \right)^k. \quad (67)$$

### 3) *Computing* $I_{\lambda,3}(w_f, \epsilon_{st})$

From (23) we recall that  $\bar{\mathbf{H}}_{i+k} = \|\mathbf{h}_{d(i-1)+2:di}\|_2, i = n-k+1, n-k+2, \dots, n$ . Then we have

$$\begin{aligned} I_{\lambda,3}(w_f, \epsilon_{st}) &= \frac{1}{\sqrt{2\pi}^{(d-1)(k)}} \int e^{-jw_f(w_{\lambda,3} - \cos^2(\epsilon_{st}) \sum_{i=n-k+1}^n \|\mathbf{h}_{d(i-1)+2:di}\|_2^2)} \prod_{i=n-k+1}^n e^{-\frac{\|\mathbf{h}_{d(i-1)+2:di}\|_2^2}{2}} d\mathbf{h}_{d(i-1)+2:di} \\ &= \frac{1}{\sqrt{2\pi}^{(d-1)(k)}} \int e^{-jw_f(\sum_{i=n-k+1}^n \bar{\mathbf{H}}_{i+k}^2 - \cos^2(\epsilon_{st}) \sum_{i=n-k+1}^n \bar{\mathbf{H}}_{i+k}^2)} \prod_{i=n-k+1}^n e^{-\frac{\bar{\mathbf{H}}_{i+k}^2}{2}} d\mathbf{h}_{d(i-1)+2:di} \\ &= \frac{1}{\sqrt{2\pi}^{(d-1)(k)}} \int \prod_{i=n-k+1}^n e^{-jw_f(\bar{\mathbf{H}}_{i+k}^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_{i+k}^2)} e^{-\frac{\bar{\mathbf{H}}_{i+k}^2}{2}} d\mathbf{h}_{d(i-1)+2:di} \\ &= \left( \frac{1}{\sqrt{2\pi}^{(d-1)}} \int e^{-jw_f(\bar{\mathbf{H}}_{i+k}^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_{i+k}^2)} e^{-\frac{\bar{\mathbf{H}}_{i+k}^2}{2}} d\mathbf{h}_{d(i-1)+2:di} \right)^k, n-k+1 \leq i \leq n. \end{aligned} \quad (68)$$

Using the distribution function for  $\bar{\mathbf{H}}_{i+k}, n-k+1 \leq i \leq n$ , from (29) we further have

$$\begin{aligned} I_{\lambda,3}(w_f, \epsilon_{st}) &= \left( \frac{1}{\sqrt{2\pi}^{(d-1)}} \int e^{-jw_f(\bar{\mathbf{H}}_{i+k}^2 - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_{i+k}^2)} e^{-\frac{\bar{\mathbf{H}}_{i+k}^2}{2}} d\mathbf{h}_{d(i-1)+2:di} \right)^k \\ &= \left( \int_{\bar{\mathbf{H}}_i \geq 0} e^{-jw_f(\bar{\mathbf{H}}_{i+k} - \cos^2(\epsilon_{st}) \bar{\mathbf{H}}_{i+k})} \frac{2^{-(d-1)/2}}{\Gamma((d-1)/2)} \bar{\mathbf{H}}_i^{(d-1)/2-1} e^{-\bar{\mathbf{H}}_i/2} d\bar{\mathbf{H}}_i \right)^k \\ &= \frac{1}{\sqrt{1 - 2j(w_f(\cos^2(\epsilon_{st}) - 1))}^{(d-1)k}}. \end{aligned} \quad (69)$$

We summarize the results from this subsection in the following theorem.

**Theorem 7.** Let  $0 < \epsilon_{st} < \frac{\pi}{2}$ . Set  $S'_w$  and  $\zeta_{\epsilon_{st}}(S'_w)$  as in (11) and (55), respectively. Let  $\sigma_{dn-1}(\cdot)$  be the standard Lebesgue measure on the unit  $S^{dn-1}$  sphere. Then

$$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})} \leq 1 - \max_{\lambda \geq 0} \lim_{\epsilon_{wf} \rightarrow 0^+} \int_{-\infty}^{\infty} \frac{1}{2\pi j(w_f - j\epsilon_{wf})} I_{\lambda,1}(w_f, \epsilon_{st}) I_{\lambda,2}(w_f, \epsilon_{st}) I_{\lambda,3}(w_f, \epsilon_{st}) dw_f, \quad (70)$$

where  $I_{\lambda,1}(w_f, \epsilon_{st})$ ,  $I_{\lambda,2}(w_f, \epsilon_{st})$ , and  $I_{\lambda,3}(w_f, \epsilon_{st})$  are as given in (66), (67), and (69), respectively.

*Proof.* Follows from the discussion presented above.  $\square$

To give an idea how close to the true values are the estimates for  $\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$  obtained through Theorem 7 we below in Tables 5 and 6 present the results for two different block lengths,  $d = 2$  and  $d = 10$ . As the tables indicate, Theorem 7 gives values that are indeed very close to the exact ones. We also add that in this introductory paper, we chose to present a simple strategy that already works fairly well. One can actually improve on it but that requires a bit of additional technical effort (similar is also true for some of the considerations from the following section). To ensure that here we preserve the elegance of the introductory concepts we will present those considerations in a separate paper.

#### 4.1.2 Determining $v_{M+2k_{st}+1}(S'_w)$

As discussed above, to determine  $v_{M+2k_{st}+1}(S'_w)$  (and ultimately  $P_{err}$  in (54)) one in principle can utilize the spherical Steiner formula (48). That would of course be possible if one can determine  $\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))$

Table 5: Simulated  $\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$  and its upper bound obtained through Theorem 7;  $d = 2, k = 6, n = 18$

$\epsilon_{st}$	0.5000	0.6000	0.7000	0.8000	0.9000
$\lambda$	0.7152	0.8441	0.9613	1.0858	1.2015
$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$ – upper bound	<b>0.1465</b>	<b>0.3538</b>	<b>0.6145</b>	<b>0.8325</b>	<b>0.9505</b>
$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$ – simulated	<b>0.1388</b>	<b>0.3410</b>	<b>0.5980</b>	<b>0.8261</b>	<b>0.9471</b>

Table 6: Simulated  $\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$  and its upper bound obtained through Theorem 7;  $d = 10, k = 6, n = 18$

$\epsilon_{st}$	0.6000	0.6500	0.7000	0.7500	0.8000
$\lambda$	1.8045	1.9119	2.0291	2.1878	2.2659
$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$ – upper bound	<b>0.0619</b>	<b>0.1877</b>	<b>0.4391</b>	<b>0.7254</b>	<b>0.9122</b>
$\frac{\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))}{\sigma_{dn-1}(S^{dn-1})}$ – simulated	<b>0.0482</b>	<b>0.1727</b>	<b>0.4346</b>	<b>0.7168</b>	<b>0.9085</b>

and if the system in (48) can be solved. To solve the system in (48) we employed the following standard  $(\lambda_{reg}, p_{reg}, q_{reg})$  regression

$$\begin{aligned}
& \min_{v_{ist}(S'_w) \geq 0} \quad \|\sigma(S'_w) - G\mathbf{v}(S'_w)\|_{p_{reg}} + \lambda_{reg} \|\mathbf{v}(S'_w)\|_{q_{reg}} \\
& \text{subject to} \quad \sigma(S'_w) - G\mathbf{v}(S'_w) \geq \mathbf{0} \\
& \quad \sum_{i_{st}=0}^{dn-2} v_{ist}(S'_w) = 1 - v_{dn-1}(S'_w) \\
& \quad \sum_{i_{st}=0}^{dn-1} (-1)^{i_{st}} v_{ist}(S'_w) = 0,
\end{aligned} \tag{71}$$

where

$$\begin{aligned}
\sigma(S'_w) &= (\sigma_{dn-1}(\zeta_{\epsilon_{st}^{(1)}}(S'_w)), \sigma_{dn-1}(\zeta_{\epsilon_{st}^{(2)}}(S'_w)), \dots, \sigma_{dn-1}(\zeta_{\epsilon_{st}^{(n_{\epsilon_{st}})}}(S'_w)))^T \\
\mathbf{v}(S'_w) &= (v_0(S'_w), v_1(S'_w), \dots, v_{dn-2}(S'_w))^T \\
G_{n_{st}, i_{st}} &= g_{n_{st}, i_{st}}, 0 \leq i_{st} \leq nd - 2, 1 \leq n_{st} \leq n_{\epsilon_{st}} \\
v_{dn-1}(S'_w) &= \frac{\sigma_{dn-1}(S'_w)}{\sigma_{dn-1}(S^{dn-1})}.
\end{aligned} \tag{72}$$

Following considerations from earlier parts of this section one then has

$$v_{dn-1}(S'_w) = \frac{\sigma_{dn-1}(S'_w)}{\sigma_{dn-1}(S^{dn-1})} = P(\sum_{i=1}^n \bar{\mathbf{H}}_i \geq 0). \tag{73}$$

Similarly to (63) and (64) we have

$$v_{dn-1}(S'_w) = \lim_{\epsilon_{wf} \rightarrow 0^+} \int_{-\infty}^{\infty} \frac{1}{2\pi j(w_f - j\epsilon_{wf})} I_{\lambda,1}^v(w_f) I_{\lambda,2}^v(w_f) dw_f, \tag{74}$$

where

$$I_{\lambda,1}^v(w_f) = \frac{1}{\sqrt{2\pi}^{d(n-k)}} \int e^{jw_f(\|\mathbf{h}_{1:d(n-k)}\|_2)} e^{-\frac{\|\mathbf{h}_{1:d(n-k)}\|_2^2}{2}} d\mathbf{h}_{1:d(n-k)}$$

$$\begin{aligned}
&= \left( \int_{\bar{\mathbf{H}}_i \geq 0} e^{jw_f \sqrt{\bar{\mathbf{H}}_i}} \frac{2^{-d/2}}{\Gamma(d/2)} \bar{\mathbf{H}}_i^{d/2-1} e^{-\bar{\mathbf{H}}_i/2} d\bar{\mathbf{H}}_i \right)^{n-k} \\
I_{\lambda,2}^v(w_f) &= \frac{1}{\sqrt{2\pi}^k} \int e^{jw_f \sum_{i=n-k+1}^n \mathbf{h}_{d(i-1)+1}} \prod_{i=n-k+1}^n e^{-\frac{\mathbf{h}_{d(i-1)+1}^2}{2}} d\mathbf{h}_{d(i-1)+1} = e^{-kw_f^2/2}.
\end{aligned} \tag{75}$$

In Figures 4 and 5 we present the results that we obtained based on the above considerations. Clearly, the figures are designed to showcase the effect that both, increasing the block-length as well as increasing the number of blocks have on the probability of error and its various estimates that we provide. We chose  $\epsilon_{st}^{(i)} \in [0.4, 1]$  and  $\epsilon_{st}^{(i)} = 0.4 + \Delta_{\epsilon_{st}} i, i \in \{1, 2, \dots\}$ . For  $d = 2, k = 6, n = 18$  we chose  $\Delta_{\epsilon_{st}} = 0.01$ , for  $d = 2, k = 24, n = 72$  we chose  $\Delta_{\epsilon_{st}} = 0.004$ , and for  $d = 10, k = 6, n = 18$  we chose  $\Delta_{\epsilon_{st}} = 0.002$ . There are several quantities that are plotted in each figure:  $p_{sim}$  which is a simulated probability of error;  $p_{num}$  which is obtained based on the above considerations with  $\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))$  simulated;  $p_{apx}$  which is obtained fully analytically based on the theoretical upper bound on  $\sigma_{dn-1}(\zeta_{\epsilon_{st}}(S'_w))$  derived in the previous section and the above regression mechanism to estimate  $\mathbf{v}(S'_w)$ . Finally we present a fully analytical upper bound on the probability of error  $p_{ub}^{(ldp)}$  obtained using results of Theorem 3 and recognizing that  $\|\mathbf{g}\|_2^2$  in Theorem 3 is a  $\chi$ -square random variable. As can be seen from figures,  $p_{sim}$  and  $p_{num}$  are almost identical (in fact they should be identical if the simulation results were fully exact and the system (48) were numerically well-posed). One can also note that analytical approximation  $p_{apx}$  is also very close to the simulated values (one can of course play with the regression parameters and the results would be a bit different; we selected  $\lambda_{reg} = 1, p_{reg} = 2$ , and  $q_{reg} = 2$ ). Finally, the fully analytical upper bound  $p_{ub}^{(ldp)}$  is in our view pretty close to the simulated values as well given how daunting task we actually had in front of us at the beginning. We should emphasize once again that the finite dimensional situation considered here is pretty much the only scenario where the high-dimensional geometry has a bit of an edge compared to the probabilistic theories that we have developed and used to obtain  $p_{ub}^{(ldp)}$ . Also, as block-length grows, the gap is even smaller and as figures indicate the transition zone is shrinking (this is of course to be expected based even just on a mere fact that the system dimensions are larger in the right side portions of the figures; moreover, this is also the key reason why we selected the smallest block length, i.e.  $d = 2$ , basically ensuring that the scenario where the gap is the largest is showcased as well).

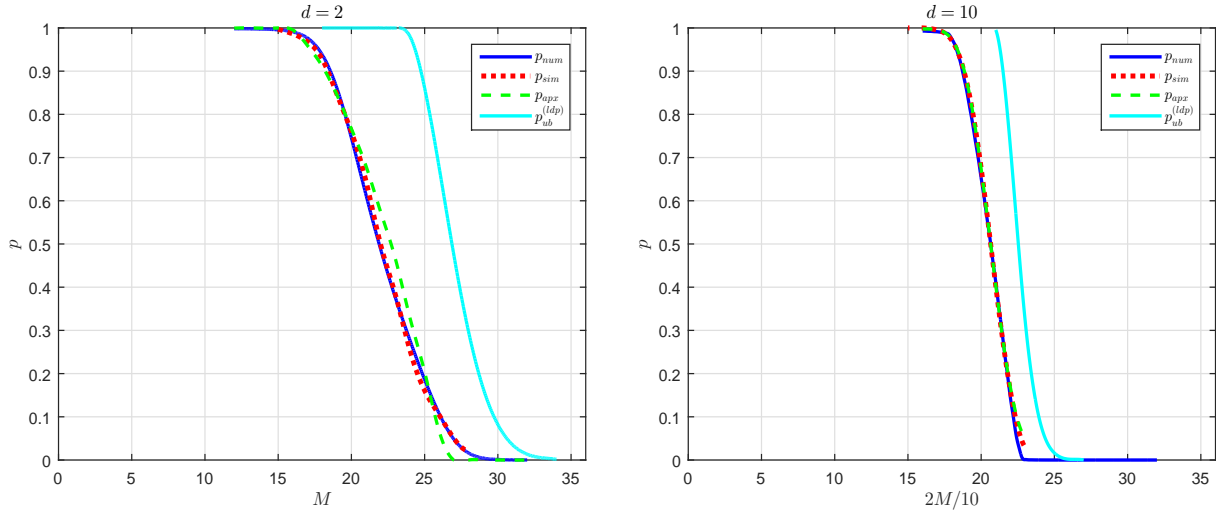


Figure 4: Probability of error estimates,  $\ell_2/\ell_1$ -optimization;  $k = 6, n = 18$

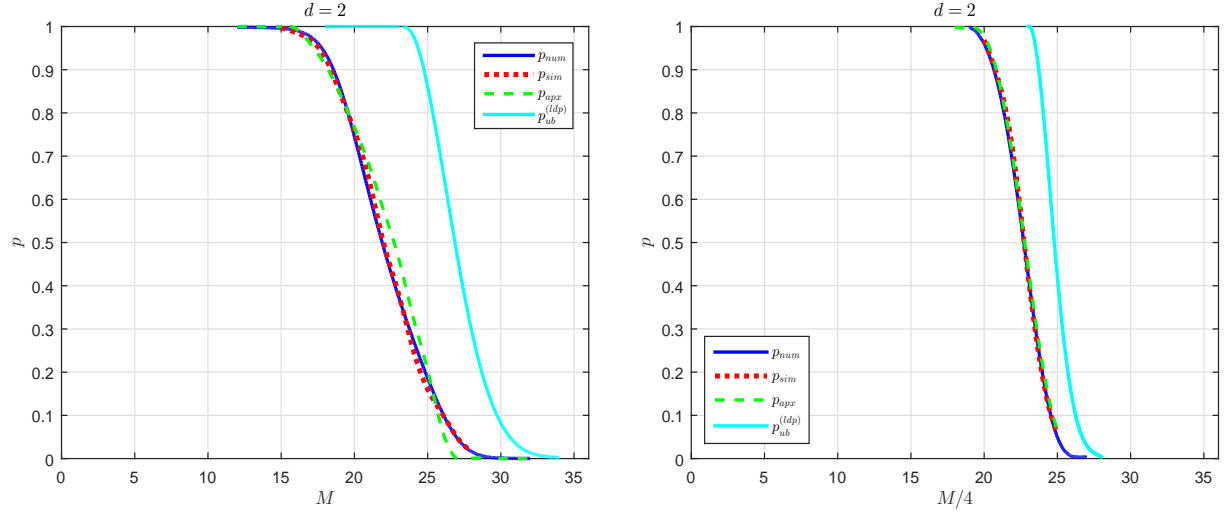


Figure 5: Probability of error estimates,  $\ell_2/\ell_1$ -optimization; left –  $k = 6$ ,  $n = 18$ ; right –  $k = 24$ ,  $n = 72$

## 4.2 Probabilistic approach

As we stated above, in Figures 4 and 5, a fully analytical upper bound on the probability of error (denoted by  $p_{ub}^{(ldp)}$  and obtained using results of Theorem 3) was plotted as well. In this subsection we will provide another way to upper-bound the probability of error. It will turn out that the results one can obtain through such an approach are similar to  $p_{ub}^{(ldp)}$ . However, the technique that we will present is of independent interest and we view it as a fairly general and useful tool.

We start things off by recalling on (12)

$$P_{err} = P(\min_{\mathbf{w} \in S'_w} \|\mathbf{A}\mathbf{w}\|_2 \leq 0) = P(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} (\mathbf{y}^T \mathbf{A}\mathbf{w}) \geq 0). \quad (76)$$

Relying on the same trick that was utilized in (35) we then have

$$P_{err} = P(\min_{\mathbf{w} \in S'_w} \|\mathbf{A}\mathbf{w}\|_2 \leq 0) = P(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} (\mathbf{y}^T \mathbf{A}\mathbf{w}) \geq 0) \leq P(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} \mathbf{y}^T \mathbf{A}\mathbf{w} + (g - t_1) \geq 0) / P(g \geq t_1). \quad (77)$$

Adjusting Lemma 2 (essentially its original form from [19]) one then has

$$\begin{aligned} P_{err} &\leq P(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} \mathbf{y}^T \mathbf{A}\mathbf{w} + (g - t_1) \geq 0) / P(g \geq t_1) \\ &\leq P(\max_{\mathbf{w} \in S'_w} \min_{\|\mathbf{y}\|_2=1} (\mathbf{y}^T \mathbf{g} + \mathbf{h}^T \mathbf{w} - t_1) \geq 0) / P(g \geq t_1) \\ &\leq P(-\|\mathbf{g}\|_2 + w(\mathbf{h}, S'_w) - t_1 \geq 0) / P(g \geq t_1) \\ &= P(-\|\mathbf{g}\|_2 + w(\mathbf{h}, S_w) - t_1 \geq 0) / P(g \geq t_1) \\ &\leq \min_{\lambda \geq 0} P(-\|\mathbf{g}\|_2 + w_\lambda(\mathbf{h}, S_w) - t_1 \geq 0) / P(g \geq t_1). \\ &\leq \min_{\lambda \geq 0} P_{\lambda, ub} / P(g \geq t_1), \end{aligned} \quad (78)$$

where clearly we set

$$P_{\lambda, ub} = P(-\|\mathbf{g}\|_2 + w_\lambda(\mathbf{h}, S_w) - t_1 \geq 0). \quad (79)$$

We then make use of the indicator step function similarly as in (61)

$$P_{\lambda, ub} = P(-\|\mathbf{g}\|_2 + w_\lambda(\mathbf{h}, S_w) - t_1 \geq 0)$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}^M} \frac{1}{\sqrt{2\pi}^{dn}} \int_{\mathbf{g}, \mathbf{h}} h(-\|\mathbf{g}\|_2 + w_\lambda(\mathbf{h}, S_w) - t_1) e^{-\frac{\|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2}{2}} d\mathbf{h} d\mathbf{g} \\
&= \frac{1}{\sqrt{2\pi}^M} \frac{1}{\sqrt{2\pi}^{dn}} \int_{\|\mathbf{g}\|_2 \geq \max(-t_1, 0), \mathbf{h}} h((w_\lambda(\mathbf{h}, S_w))^2 - (t_1 + \|\mathbf{g}\|_2)^2) e^{-\frac{\|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2}{2}} d\mathbf{h} d\mathbf{g} \\
&\quad + \frac{1}{\sqrt{2\pi}^M} \int_{\|\mathbf{g}\|_2 \leq \max(-t_1, 0)} e^{-\frac{\|\mathbf{g}\|_2^2}{2}} d\mathbf{g} \\
&= P_{\lambda, ub}^{(1)} + \left( \int_{0 \leq z_g \leq (\max(-t_1, 0))^2} \frac{2^{-M/2}}{\Gamma(M/2)} z_g^{M/2-1} e^{-z_g/2} dz_g \right), \tag{80}
\end{aligned}$$

where

$$P_{\lambda, ub}^{(1)} = \frac{1}{\sqrt{2\pi}^M} \frac{1}{\sqrt{2\pi}^{dn}} \int_{\|\mathbf{g}\|_2 \geq \max(-t_1, 0), \mathbf{h}} h((w_\lambda(\mathbf{h}, S_w))^2 - (t_1 + \|\mathbf{g}\|_2)^2) e^{-\frac{\|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2}{2}} d\mathbf{h} d\mathbf{g}. \tag{81}$$

Combining (62), (80), and (81) we then have

$$\begin{aligned}
P_{\lambda, ub}^{(1)} &= \frac{1}{\sqrt{2\pi}^M} \frac{1}{\sqrt{2\pi}^{dn}} \int_{\|\mathbf{g}\|_2 \geq \max(-t_1, 0), \mathbf{h}} h((w_\lambda(\mathbf{h}, S_w))^2 - (t_1 + \|\mathbf{g}\|_2)^2) e^{-\frac{\|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2}{2}} d\mathbf{h} d\mathbf{g} \\
&= \frac{1}{\sqrt{2\pi}^M} \frac{1}{\sqrt{2\pi}^{dn}} \int_{\|\mathbf{g}\|_2 \geq \max(-t_1, 0), \mathbf{h}} \lim_{\epsilon_{wf} \rightarrow 0+} \int_{-\infty}^{\infty} \frac{e^{jw_f((w_\lambda(\mathbf{h}, S_w))^2 - (t_1 + \|\mathbf{g}\|_2)^2)}}{2\pi j(w_f - j\epsilon_{wf})} dw_f e^{-\frac{\|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2}{2}} d\mathbf{h} d\mathbf{g} \\
&= \lim_{\epsilon_{wf} \rightarrow 0+} \int_{-\infty}^{\infty} \frac{1}{2\pi j(w_f - j\epsilon_{wf})} \int_{\|\mathbf{g}\|_2 \geq \max(-t_1, 0), \mathbf{h}} e^{jw_f(w_{\lambda,1} + w_{\lambda,2} + w_{\lambda,3} - (t_1 + \|\mathbf{g}\|_2)^2)} \frac{e^{-\frac{\|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2}{2}}}{\sqrt{2\pi}^{M+dn}} d\mathbf{h} d\mathbf{g} dw_f \\
&= \lim_{\epsilon_{wf} \rightarrow 0+} \int_{-\infty}^{\infty} \frac{1}{2\pi j(w_f - j\epsilon_{wf})} I_{\lambda,1}(-w_f, \pi/2) I_{\lambda,2}(-w_f, \pi/2) I_{\lambda,3}(-w_f, \pi/2) I_{\lambda,4}(w_f, t_1) dw_f, \tag{82}
\end{aligned}$$

where

$$\begin{aligned}
I_{\lambda,4}(w_f, t_1) &= \frac{1}{\sqrt{2\pi}^M} \int_{\|\mathbf{g}\|_2 \geq \max(-t_1, 0)} e^{-jw_f(t_1 + \|\mathbf{g}\|_2)^2} e^{-\frac{\|\mathbf{g}\|_2^2}{2}} d\mathbf{g} \\
&= \left( \int_{z_g \geq (\max(-t_1, 0))^2} e^{-jw_f(t_1 + \sqrt{z_g})^2} \frac{2^{-M/2}}{\Gamma(M/2)} z_g^{M/2-1} e^{-z_g/2} dz_g \right). \tag{83}
\end{aligned}$$

We summarize the above discussion in the following theorem.

**Theorem 8.** Assume the setup of Theorem 3. Accordingly, let  $P_{err}$  be the probability that the solution of (5) is not the  $k$ -block-sparse solution of (2). Then

$$\begin{aligned}
P_{err} &\leq \min_{t_1} \min_{\lambda \geq 0} \left( \lim_{\epsilon_{wf} \rightarrow 0+} \int_{-\infty}^{\infty} \frac{1}{2\pi j(w_f - j\epsilon_{wf})} I_{\lambda,1}(-w_f, \pi/2) I_{\lambda,2}(-w_f, \pi/2) I_{\lambda,3}(-w_f, \pi/2) I_{\lambda,4}(w_f, t_1) dw_f \right. \\
&\quad \left. + \left( \int_{0 \leq z_g \leq (\max(-t_1, 0))^2} \frac{2^{-M/2}}{\Gamma(M/2)} z_g^{M/2-1} e^{-z_g/2} dz_g \right) / P(g \geq t_1) \right), \tag{84}
\end{aligned}$$

where  $I_{\lambda,1}(\cdot, \cdot)$ ,  $I_{\lambda,2}(\cdot, \cdot)$ ,  $I_{\lambda,3}(\cdot, \cdot)$ , and  $I_{\lambda,4}(\cdot, \cdot)$  are as given in (66), (67), (69), and (83) respectively.

*Proof.* Follows from the previous discussion.  $\square$

In Figure 6 we present the results that one can obtain based on the above theorem. Quantity denoted by  $p_{ub}^{(ag)}$  is the upper bound on  $P_{err}$  that we computed utilizing Theorem 8. As can be seen (and as we hinted at the beginning of this subsection) the results are comparable to  $p_{ub}^{(ldp)}$  (as expected this is even more emphasized as the system dimensions grow). However, the strategy that we developed above is a very powerful tool and we wanted to present it on its own without too much emphasis on the concrete values that it provides for the upper estimate of  $P_{err}$ .

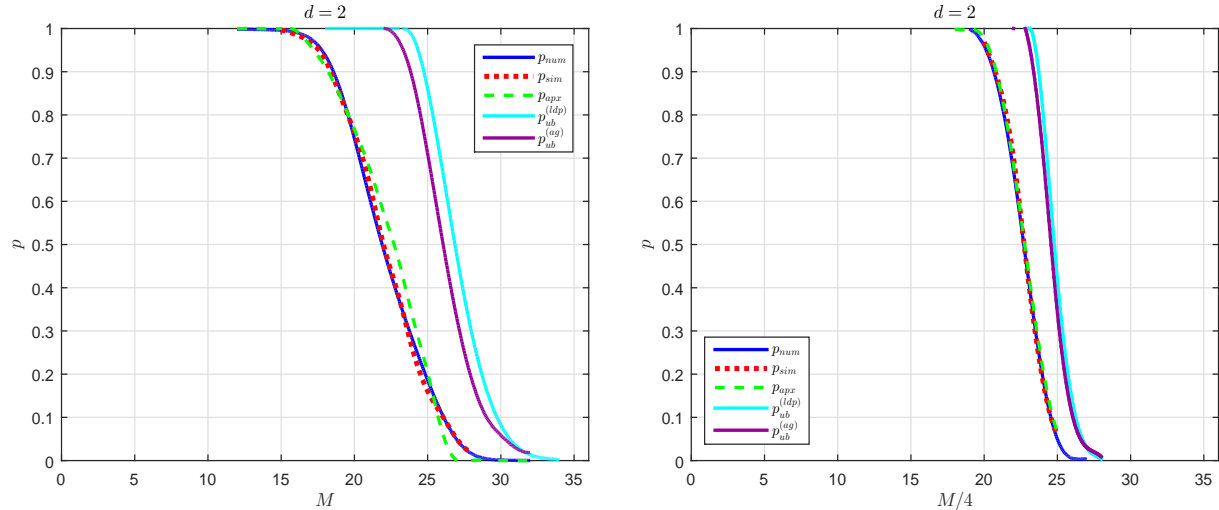


Figure 6: Probability of error estimates,  $\ell_2/\ell_1$ -optimization; left –  $k = 6$ ,  $n = 18$ ; right –  $k = 24$ ,  $n = 72$

## 5 Conclusion

The main topic of interest in this paper are random under-determined systems of linear equations with block-sparse solutions. A particular line of study that focuses on a special class of the so called  $\ell_2/\ell_1$ -optimization algorithms was considered. We first revisited a few earlier results that relate to an asymptotic analysis of such algorithms when employed for solving random linear systems with block-sparse solutions. In particular, we first revisited the so-called phase-transition phenomena. These phenomena provide the exact characterization of the so-called breaking points, i.e. of the points where the algorithms sharply transition from being perfectly reliable to being unable to solve problems at hand. In a series of earlier works we introduced a systematic way for studying such a phenomena that relies on a purely probabilistic approach.

We then went a bit further and created a completely new concept that we in a way connected to the large deviation theory. Namely, we introduced a large deviation type of considerations for the probabilities of error which in essence deal with the tail behavior of the phase-transition curves. These are again by their nature asymptotic and we were able to show that the mechanisms we designed earlier can in fact be restructured so that they can handle such a type of problems as well.

In addition to the asymptotic considerations, we also raised the level a bit and considered finite dimensional scenarios as well (of course these being the ultimate considerations one would like to have). We then showed that the probabilistic theories that we have built can in fact produce a solid finite dimensional estimates as well.

Finally, we also considered a high-dimensional integral geometry approach as an alternative tool for the types of the analyses of interest in this paper. We introduced several novel technical steps that enable such an approach to work which eventually enabled us to obtain excellent estimates for success/failure of the  $\ell_2/\ell_1$ -optimization when used for solving linear systems with block-sparse solutions.

Of course, as is often the case with the type of results that we showcased here, one can then continue further and consider various other aspects/extensions of the algorithms/problems at hand and adjust the techniques introduced here and in a few of our earlier works so that they fit those problems as well. Such adjustments are fairly standard and for a few particularly interesting problems we will in several companion papers present how one can do them and what kind of results one can eventually obtain through them.

## References

- [1] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde. Model-based compressive sensing. available online at <http://www.dsp.ece.rice.edu/cs/>.
- [2] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and Richard Baraniuk. Distributed compressed sensing. *Allerton*, 2005.
- [3] T. Blumensath and M. E. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Transactions on Information Theory*, 55(4):187–1882, 2009.
- [4] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52(12):489–509, 2006.
- [5] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk. Recovery of clustered sparse signals from compressive measurements. *SAMPTA, International Conference on Sampling Theory and Applications*, 2009. Marseille, France.
- [6] J. Chen and X. Huo. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Trans. on Signal Processing*, 54(12):4634–4643, 2006.
- [7] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. on Signal Processing*, 53(7):2477 – 2488, 2005.
- [8] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. available online at <http://arxiv.org/abs/0803.0811>.
- [9] D. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. 2004. Technical report, Department of Statistics, Stanford University.
- [10] D. Donoho. High-dimensional centrally symmetric polytopes with neighborlines proportional to dimension. *Disc. Comput. Geometry*, 35(4):617–652, 2006.
- [11] D. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. *Proc. National Academy of Sciences*, 106(45):18914–18919, 2009.
- [12] D. L. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52(4):1289–1306, 2006.
- [13] D. L. Donoho, Y. Tsaig, I. Drori, and J.L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. 2007. available online at <http://www.dsp.ece.rice.edu/cs/>.
- [14] Y. C. Eldar and H. Bolcskei. Block-sparsity: Coherence and efficient recovery. *ICASSP, IEEE International Conference on Acoustics, Signal and Speech Processing*, pages 2885–2888, April 2009. Taipei, Taiwan.
- [15] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. Compressed sensing of block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. on Signal Processing*, 58(6):3042–3054, 2010.
- [16] Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. on Information Theory*, 55(1):5302–5316, 2009.
- [17] Y. C. Eldar and H. Rauhut. Average case analysis of multichannel sparse recovery using convex relaxation. *IEEE Trans. on Information Theory*, 56(1):505–519, 2010.
- [18] A. Ganesh, Z. Zhou, and Y. Ma. Separation of a subspace-sparse signal: Algorithms and conditions. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3141–3144, April 2009. Taipei, Taiwan.
- [19] Y. Gordon. Some inequalities for gaussian processes and applications. *Israel Journal of Mathematics*, 50(4):265–289, 1985.

- [20] M. Mishali and Y. Eldar. Reduce and boost: Recovering arbitrary sets of jointly sparse vectors. *IEEE Trans. on Signal Processing*, 56(10):4692–4702, 2008.
- [21] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.
- [22] D. Needell and R. Vershynin. Unifrom uncertainly principles and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9(3):317–334, 2009.
- [23] S. Negahban and M. J. Wainwright. Simultaneous support recovery in high dimensions: Benefits and perils of block  $\ell_1/\ell_\infty$ -regularization. *IEEE Trans. on Information Theory*, 57(6):3841–3863, 2011.
- [24] F. Parvaresh and B. Hassibi. Explicit measurements with almost optimal thresholds for compressed sensing. *ICASSP, IEEE International Conference on Acoustics, Signal and Speech Processing*, Mar-Apr 2008. Las Vegas, NV.
- [25] R. Schneider and W. Weil. *High-dimensional integral geometry*. Springer, 2008.
- [26] M. Stojnic. Block-length dependent thresholds in block-sparse compressed sensing. available online at <http://arxiv.org/abs/0907.3679>.
- [27] M. Stojnic. Compressed sensing of block-sparse positive vectors. available online at <http://arxiv.org/abs/1306.3977>.
- [28] M. Stojnic. Lifting/lowering Hopfield models ground state energies. available online at <http://arxiv.org/abs/1306.3975>.
- [29] M. Stojnic. Linear under-determined systems with sparse solutions: Redirecting a challenge? available online at <http://arxiv.org/abs/1304.0004>.
- [30] M. Stojnic. Optimality of  $\ell_2/\ell_1$ -optimization block-length dependent thresholds. available online at <http://arxiv.org/abs/1304.0001>.
- [31] M. Stojnic. Upper-bounding  $\ell_1$ -optimization weak thresholds. available online at <http://arxiv.org/abs/1303.7289>.
- [32] M. Stojnic. Various thresholds for  $\ell_1$ -optimization in compressed sensing. available online at <http://arxiv.org/abs/0907.3666>.
- [33] M. Stojnic. A simple performance analysis of  $\ell_1$ -optimization in compressed sensing. *ICASSP, IEEE International Conference on Acoustics, Signal and Speech Processing*, pages 3021–3024, April 2009. Taipei, Taiwan.
- [34] M. Stojnic. Strong thresholds for  $\ell_2/\ell_1$ -optimization in block-sparse compressed sensing. *ICASSP, IEEE International Conference on Acoustics, Signal and Speech Processing*, pages 3025–3028, April 2009. Taipei, Taiwan.
- [35] M. Stojnic.  $\ell_2/\ell_1$ -optimization in block-sparse compressed sensing and its strong thresholds. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):350–357, 2010.
- [36] M. Stojnic. Recovery thresholds for  $\ell_1$  optimization in binary compressed sensing. *ISIT, IEEE International Symposium on Information Theory*, pages 1593 – 1597, 13-18 June 2010. Austin, TX.
- [37] M. Stojnic. Towards improving  $\ell_1$  optimization in compressed sensing. *ICASSP, IEEE International Conference on Acoustics, Signal and Speech Processing*, pages 3938–3941, 14-19 March 2010. Dallas, TX.
- [38] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. on Signal Processing*, 57(8):3075–3085, 2009.
- [39] V.N. Temlyakov. A remark on simultaneous greedy approximation. *East J. Approx.*, 100, 2004.



- [40] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. on Information Theory*, 53(12):4655–4666, 2007.
- [41] J. Tropp, A. C. Gilbert, and M. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.
- [42] J. A. Tropp. Greed is good: algorithmic results for sparse approximations. *IEEE Trans. on Information Theory*, 50(10):2231–2242, 2004.
- [43] E. van den Berg and M. P. Friedlander. Joint-sparse recovery from multiple measurements. available online at <https://arxiv.org/abs/0904.2051>.
- [44] H. Vikalo, F. Parvaresh, and B. Hassibi. On sparse recovery of compressed dna microarrays. *Asilomar conference*, November 2007.
- [45] A. C. Zelinski, V. K. Goyal, and E. Adalsteinsson. Simultaneously sparse solutions to linear inverse problems with multiple system matrices and a single observation vector. *SIAM J. Sci. Comput.*, 31(6):4533–4579, 2010.
- [46] A. C. Zelinski, L. L. Wald, K. Setsompop, V. K. Goyal, and E. Adalsteinsson. Sparsity-enforced slice-selective mri rf excitation pulse design. *IEEE Trans. on Medical Imaging*, 27(9):1213–1229, 2008.